

DIE GRAPHISCHE DARSTELLUNG DER DON'T KNOW-KATEGORIE IN RATINGSKALEN

Inaugural dissertation submitted in partial fulfillment of the requirements for
the degree Doctor of Social Sciences in the Graduate School of Economic and
Social Sciences at the University of Mannheim

Kathrin Bogner, geb. Wohn

Dekan: Prof. Dr. Michael Diehl

Betreuer / Erstgutachter: Prof. Dr. Peter Ph. Mohler

Betreuer / Zweitgutachter: Prof. Dr. Jürgen H.P. Hoffmeyer-Zlotnik

Drittgutachterin: Prof. Dr. Annelies G. Blom

Tag der Disputation: 29.April 2015

DANKSAGUNG

Eine wissenschaftliche Arbeit ist nie das Werk einer einzelnen Person, deshalb ist es jetzt an der Zeit, mich bei allen Menschen zu bedanken, die mir die Erstellung meiner Dissertation ermöglicht haben. Zu besonderem Dank bin ich meinen Professoren verpflichtet. Als erster Gutachter hat mich Herr Professor Mohler stets mit seinen Anregungen dazu unterstützt, meine Arbeit weiterzuentwickeln, dabei stets offen für die Erkenntnisse verschiedener Disziplinen zu sein und trotzdem das Ziel nicht aus den Augen zu verlieren. Hätte er nicht mein Interesse an der empirischen Sozialforschung und ihren Methoden geweckt, wäre diese Arbeit nie zustande gekommen. Auch meinem Zweitgutachter Herr Professor Hoffmeyer-Zlotnik möchte ich an dieser Stelle dafür danken, dass er mich so intensiv mit seinem akademischen Rat auf diesem Weg begleitet hat. Schließlich bin ich auch Frau Professorin Blom für ihr drittes Gutachten zu Dank verpflichtet.

Ebenso geht mein Dank an meine ehemaligen Kommilitonen und meine Kollegen, die mich in den vergangenen Jahren mit bereichernden Tipps und Diskussionsbeiträgen wiederholt in neue thematische Bahnen gelenkt haben. Dabei danke ich besonders Natalja Menold dafür, dass mein erstes Experiment in ihrem Projekt erhoben werden konnte. Auch ihre fachlichen Ratschläge, ihre Hilfestellung und ihr „offenes Ohr“ haben sehr viel zum Gelingen dieser Arbeit beigetragen. Klaus Pforr gilt mein großer Dank für die Begleitung der Entwicklung des statistischen Modells und der Graphiken, die den Ergebnissen dieser Arbeit erst ihre Aussagekraft verleihen.

Der Universität Mannheim bin ich in zweierlei Hinsicht zu Dank verpflichtet: Erstens dafür, dass sie mich in die Graduate School of Economic and Social Sciences aufgenommen hat. Dort habe ich die Möglichkeit erfahren, meine Fähigkeiten entsprechend meiner Interessen weiterzuentwickeln. Zweitens danke ich der Universität Mannheim für die Vergabe des Brigitte Schlieben-Lange-Stipendiums für Nachwuchswissenschaftlerinnen mit Kind. Mit dieser finanziellen Unterstützung war es mir möglich, die Arbeit an der Promotion schnell nach unserem Familienzuwachs wieder aufzunehmen.

Schließlich möchte ich mich bei meiner Familie und meinen Freunden bedanken. Meinem Mann danke ich dafür, dass er mir die Freiräume verschafft hat, mich dieser Arbeit widmen zu können, aber auch dafür, dass er mich hin und wieder mit schönen Dingen erfolgreich von ihr abgelenkt hat. Meinen Eltern bin ich zu großem Dank verpflichtet, dass sie mich mein Leben lang dabei unterstützt haben, meinen eigenen Weg zu finden und auch zu gehen. Nun selbst Mutter, weiß ich, dass dies gar nicht so einfach ist. Außerdem gilt mein herzlicher Dank meinen Schwiegereltern, ohne deren Unterstützung vieles nicht möglich gewesen wäre.

INHALTSVERZEICHNIS

Abkürzungsverzeichnis.....	XI
Abbildungsverzeichnis.....	XII
Tabellenverzeichnis	XVI
I EINLEITUNG.....	1
II THEORIE & STAND DER FORSCHUNG.....	7
1. Bedeutung des Layouts von Ratingskalen.....	7
1.1 Verbale und nonverbale Ratingskalenelemente	7
1.2 Ratingskalenelemente im konversationalen Kontext von Umfragen	9
1.3 Heuristiken der Interpretation nonverbaler Ratingskalenelemente	11
2. Effekte des Layouts von Ratingskalen.....	13
2.1 Stand der Forschung	13
2.2 Kovariate von Layouteffekten	15
2.2.1 Eigenschaften der Befragten.....	16
2.2.2 Eigenschaften der Ratingskala.....	20
3. Don't know-Kategorie in Ratingskalen	21
3.1 Diskussion der Don't know-Kategorie in Ratingskalen.....	21
3.2 Survey Satisficing-Theorie	24
3.3 Don't know-Vorgabe in verschiedenen Erhebungsmodi	25
3.4 Kovariate von Don't know-Antworten.....	27
3.5 Empfehlungen für die Praxis	27
4. Effekte des Layouts der Don't know-Kategorie in Ratingskalen	29
4.1 Die Studie von Tourangeau, Couper und Conrad	30
4.2 Die Studie von Christian, Parsons und Dillman	32
4.3 Mögliche Ursachen der konträren Ergebnisse beider Studien	33

5.	Forschungsfragen & Hypothesen.....	36
5.1	Studie 1	37
5.2	Studie 2	40
5.3	Hypothesen	41
III	STUDIE 1	45
6.	Studiendesign	45
6.1	Daten	46
6.2	Items	46
6.3	Methode	50
6.3.1	Auswahl der Analyse­methode	50
6.3.2	Logistische Multilevel-Regression mit fixen Effekten.....	52
6.3.3	Interpretation der Ergebnisse anhand marginaler Effekte	55
6.3.4	Aufbau der logistischen Regressionsmodelle.....	59
7.	Deskriptive Ergebnisse.....	60
7.1	Antwortverteilung der Political Efficacy-Skala	61
7.2	Antwortverteilung der Items zur Erfassung von Lernstrategien	62
7.3	Antwortverteilungen der Items und des Indexes der Einstellungsstärke.....	63
7.4	Extreme, mittlere und Don't Know-Antworten in den Item-Konstrukten	65
8.	Bivariate Ergebnisse.....	67
8.1	Kategoriewahl in Abhängigkeit von Skalenlayout.....	68
8.2	Kategoriewahl in Abhängigkeit von Skalenlayout und Einstellungsstärke ...	72
9.	Ergebnisse der logistischen Multilevel-Regressionen.....	80
9.1	Wahl der rechten Extremkategorie	80
9.1.1	Modell 1: Effekt des Skalenlayouts	80
9.1.2	Modell 2: Effekt des Item-Konstrukts	83
9.1.3	Modell 3: Interaktionseffekt „Item-Konstrukt \times Layout“	84
9.1.4	Modell 4: Effekt der Einstellungsstärke	86
9.1.5	Modell 5: Interaktionseffekt „Einstellungsstärke \times Layout“	88
9.1.6	Zwischenfazit.....	94
9.2	Wahl der Mittelkategorie.....	102
9.2.1	Modell 1: Effekt des Layouts	102
9.2.2	Modell 2: Effekt des Item-Konstrukts	104
9.2.3	Modell 3: Interaktionseffekt „Item-Konstrukt \times Layout“	105
9.2.4	Modell 4: Effekt der Einstellungsstärke	107
9.2.5	Modell 5: Interaktionseffekt „Einstellungsstärke \times Layout“	109
9.2.6	Zwischenfazit.....	115

9.3	Wahl der DK-Kategorie	123
9.3.1	Modell 1: Effekt des Layouts	123
9.3.2	Modell 2: Effekt des Item-Konstrukts	124
9.3.3	Modell 3: Interaktionseffekt „Item-Konstrukt \times Layout“	125
9.3.4	Modell 4: Effekt der Einstellungsstärke	127
9.3.5	Modell 5: Interaktionseffekt „Einstellungsstärke \times Layout“	128
9.3.6	Zwischenfazit.....	132
10.	Fazit der Studie 1.....	139
IV	STUDIE 2	144
11.	Effekte nonverbaler Elemente in Mail und Web Survey	144
12.	Studiendesign	148
12.1	Daten	149
12.2	Items	150
12.3	Methode.....	152
12.4	Aufbau der logistischen Regressionsmodelle.....	152
13.	Deskriptive Ergebnisse.....	153
13.1	Antwortverteilung der EU-Items	153
13.2	Extreme, mittlere und Don't Know-Antworten zu den EU-Items	154
14.	Bivariate Ergebnisse der Kategoriewahl in Abhängigkeit von Skalenlayout.....	156
15.	Ergebnisse der logistischen Multilevel-Regressionen.....	158
15.1	Modell 1: Effekte des Layouts	159
15.1.1	Wahl der rechten Extremkategorie	159
15.1.2	Wahl der Mittelkategorie	161
15.1.3	Wahl der Don't Know-Kategorie	163
15.1.4	Zwischenfazit.....	164
15.2	Modell 2a: Einfluss der Bildung.....	168
15.2.1	Wahl der rechten Extremkategorie	168
15.2.2	Wahl der Mittelkategorie	169
15.2.3	Wahl der Don't Know-Kategorie	170
15.2.4	Zwischenfazit.....	171
15.3	Modell 2b: Einfluss des Alters	175
15.3.1	Wahl der rechten Extremkategorie	175
15.3.2	Wahl der Mittelkategorie	176
15.3.3	Wahl der Don't Know-Kategorie	177
15.3.4	Zwischenfazit.....	177
16.	Fazit der Studie 2.....	180

V	ZUSAMMENFASSENDE SCHLUSSBETRACHTUNG	185
	Literaturverzeichnis	198
	Eidesstattliche Erklärung	208
	Anhang	i

ABKÜRZUNGSVERZEICHNIS

<i>ADD-DK-Layout</i>	Ratingskalenlayout mit einer graphisch nicht abgetrennten, nicht-inhaltlichen Antwortkategorie
<i>DK-Kategorie</i>	Don't know-Kategorie; Synonym für nicht-inhaltliche Antwortkategorien, unabhängig deren verbalen Etikettierung
<i>DK-Antwort</i>	Don't know-Antwort; Synonym für eine Antwortabgabe anhand einer nicht-inhaltlichen Antwortkategorie
<i>LIST</i>	Inventar zur Erfassung von Lernstrategien im Studium
<i>LIST-Items</i>	Items des LIST-Inventars
<i>NO-DK-Layout</i>	Ratingskalenlayout ohne eine nicht-inhaltliche Antwortkategorie
<i>PES</i>	Skala zur Erfassung des Konzepts der politischen Selbstwirksamkeit, im Englischen: Political Efficacy Scale
<i>PES-Items</i>	Items der PES
<i>SEP-DK-Layout</i>	Ratingskalenlayout mit einer graphisch abgetrennten, nicht-inhaltlichen Antwortkategorie

ABBILDUNGSVERZEICHNIS

Abbildung 1 Drei Varianten der graphischen Darstellung der DK-Kategorie einer Ratingskala.....	30
Abbildung 2 SEP-DK- und ADD-DK-Layout der Ratingskalen bei Tourangeau et al. (2004) und Christian et al. (2009).....	33
Abbildung 3 Linksschiefe Antwortverteilung des Item 1 bei Tourangeau et al. (2004).....	34
Abbildung 4 Rechtsschiefe Antwortverteilung des Item 2	35
Abbildung 5 Experimentell variierte Skalenlayouts der Studie 1 und Studie 2	38
Abbildung 6 Multilevel-Modell der Gruppierung von experimentell variierten Items in Befragten und den Level-1 und Level-2 Variablen der Studie 1.....	51
Abbildung 7 Datensatz im „long-format“ (Items = Fälle).....	54
Abbildung 8 Graphik der marginalen Effekte (AMEs) von Befragten im NO-DK-Layout nach Einstellungsstärke (Referenz: ADD-DK).....	58
Abbildung 9 Aufbau der geschachtelten logistischen Multilevel-Regressionen der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie	60
Abbildung 10 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK)	82
Abbildung 11 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout	82
Abbildung 12 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Item-Konstrukt.....	84
Abbildung 13 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout und Item-Konstrukt	85
Abbildung 14 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Einstellungsstärke	87

Abbildung 15 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Einstellungsstärke	88
Abbildung 16 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout und Einstellungsstärke.....	89
Abbildung 17 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK), Einstellungsstärke und Item-Konstrukt.....	91
Abbildung 18 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout, Einstellungsstärke und Item-Konstrukt	93
Abbildung 19 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK)	103
Abbildung 20 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout.....	103
Abbildung 21 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Item-Konstrukt.....	106
Abbildung 22 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout und Item-Konstrukt	107
Abbildung 23 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Einstellungsstärke.....	108
Abbildung 24 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Einstellungsstärke	110
Abbildung 25 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout und Einstellungsstärke	111
Abbildung 26 Marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK), Einstellungsstärke und Item-Konstrukt	112
Abbildung 27 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout, Einstellungsstärke und Item-Konstrukt.....	114
Abbildung 28 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout.....	124
Abbildung 29 Durchschnittliche marginale Effekte (AMEs) der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Item-Konstrukt	125
Abbildung 30 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout und Item-Konstrukt	126
Abbildung 31 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Einstellungsstärke.....	127

Abbildung 32 Durchschnittliche marginale Effekte (AMEs) auf die Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Einstellungsstärke.....	129
Abbildung 33 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout und Einstellungsstärke	129
Abbildung 34 Durchschnittliche marginale Effekte (AMEs) der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK), Einstellungsstärke und Item-Konstrukt.....	130
Abbildung 35 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout, Einstellungsstärke und Item-Konstrukt.....	131
Abbildung 36 Screenshot der Single-Item-Darstellung des ersten EU-Items in der Online-Befragung der Studie 2	151
Abbildung 37 Aufbau der geschachtelten logistischen Multilevel-Regressionen der Wahl der rechten Extrem-, der Mittel- bzw. der DK-Kategorie.....	153
Abbildung 38 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK)	160
Abbildung 39 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie	160
Abbildung 40 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK)	161
Abbildung 41 Vorhergesagte Wahrscheinlichkeit der Wahl der Mittelkategorie nach Layout	162
Abbildung 42 Durchschnittlicher marginaler Effekte (AME) der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK)	163
Abbildung 43 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Bildung.....	169
Abbildung 44 Durchschnittliche marginale Effekt (AMEs) auf die Wahrscheinlichkeit der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Bildung.....	170
Abbildung 45 Durchschnittliche marginale Effekte (AMEs) auf Wahrscheinlichkeit der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Bildung.....	171
Abbildung 46 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Altersgruppe	175

Abbildung 47 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Altersgruppe	176
Abbildung 48 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Altersgruppe	177
Abbildung 49 Übersicht über die Unterstützung der Hypothesen hinsichtlich des Einflusses des Skalenlayouts auf das Antwortverhalten der Ergebnisse der Studien 1 und 2 und deren Übereinstimmung	183
Abbildung 50 Vorhergesagte Wahrscheinlichkeit der Wahl der rechten Extremkategorie nach Layout in Studie 1 (links) und Studie 2 (rechts)	187
Abbildung 51 Unterschiede der Skalenlayouts in der Darstellung der rechten Extremkategorie	187
Abbildung 52 Wahl der Mittelkategorie nach Layout in Studie 1 (links) und Studie 2 (rechts)	188
Abbildung 53 Wahl der DK-Kategorie nach Layout (links) und nach Layout und Item-Konstrukt (rechts) in Studie 1	189

TABELLENVERZEICHNIS

TABELLEN DES HAUPTTEILS

Tabelle 1 Antwortverteilungen des Items Sicherheit bei der Beantwortung der Items zur politischen Selbstwirksamkeit (PES) und zu den eigenen Lernstrategien (LIST)	64
Tabelle 2 Antwortverteilungen des Items „Wichtigkeit“ des Themas der politischen Selbstwirksamkeit (PES) und der eigenen Lernstrategien (LIST).....	64
Tabelle 3 Auswahlhäufigkeiten der rechten Extrem-, Mittel- und DK-Kategorie in der PES.....	65
Tabelle 4 Auswahlhäufigkeiten der rechten Extrem-, Mittel- und DK-Kategorie für LIST	67
Tabelle 5 Wahl der rechten Extrem-, Mittel- und DK-Kategorie in den Item-Konstrukten nach Skalenlayout	69
Tabelle 6 Wahl der rechten Extrem-, Mittel- und DK-Kategorie in den Item-Konstrukten nach Skalenlayout und Einstellungsstärke	75
Tabelle 7 Ergebnisse der logistischen Multilevel-Modelle zur Wahl der rechten Extremkategorie	98
Tabelle 8 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie.....	99
Tabelle 9 Vorhergesagte Wahrscheinlichkeiten (Whs.) der Wahl der rechten Extremkategorie	100
Tabelle 10 Ergebnisse der logistischen Multilevel-Modelle zur Wahl der Mittelkategorie	119
Tabelle 11 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der Mittelkategorie	120
Tabelle 12 Vorhergesagte Wahrscheinlichkeiten (Whs.) der Wahl der Mittelkategorie	121

Tabelle 13 Ergebnisse der logistischen Multilevel-Modelle zur Wahl der DK-Kategorie	135
Tabelle 14 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der DK-Kategorie	136
Tabelle 15 Vorhergesagte Wahrscheinlichkeiten (Whs.) der Wahl der DK-Kategorie	137
Tabelle 16 Auswahlhäufigkeiten der rechten Extrem-, Mittel- und DK-Kategorie für EU-Items	155
Tabelle 17 Wahl der rechten Extrem-, Mittel- und DK-Kategorie zu den EU-Items nach Skalenlayout	157
Tabelle 18 Ergebnisse der logistischen Multilevelmodelle zur Wahl (a) der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie	166
Tabelle 19 Durchschnittliche Marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl (a) der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie	167
Tabelle 20 Vorhergesagte Wahrscheinlichkeiten (Whs.) (a) der Wahl der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie	167
Tabelle 21 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl (a) der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie nach Layout (Referenz: ADD-DK) und Bildung.....	174
Tabelle 22 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl (a) der rechten Extremkategorie, (b) der Mittelkategorie und (c) der DK-Kategorie nach Layout (Referenz: ADD-DK) und Altersgruppe	179

TABELLEN DES ANHANGS

Tabelle A 1 Verteilung des Item 1 der PES.....	i
Tabelle A 2 Verteilung des Item 2 der PES.....	i
Tabelle A 3 Verteilung des Item 3 der PES.....	ii
Tabelle A 4 Verteilung des Item 4 der PES.....	ii
Tabelle A 5 Verteilung des Item 5 der PES.....	iii
Tabelle A 6 Verteilung des Item 6 der PES.....	iii
Tabelle A 7 Verteilung des Item 7 der PES.....	iv
Tabelle A 8 Deskriptive Statistik des summierten Index der PES-Items	iv

Tabelle A 9 Ergebnisse der Varianzanalyse der PES-Items	v
Tabelle A 10 Verteilung des Item 1 des LIST-Inventars.....	v
Tabelle A 11 Verteilung des Item 2 des LIST-Inventars	vi
Tabelle A 12 Verteilung des Item 3 des LIST-Inventars.....	vi
Tabelle A 13 Verteilung des Item 4 des LIST-Inventars.....	vii
Tabelle A 14 Verteilung des Item 5 des LIST-Inventars.....	vii
Tabelle A 15 Verteilung des Item 6 des LIST-Inventars.....	viii
Tabelle A 16 Verteilung des Item 7 des LIST-Inventars.....	viii
Tabelle A 17 Verteilung des Item 8 des LIST-Inventars.....	ix
Tabelle A 18 Verteilung des Item 9 des LIST-Inventars.....	ix
Tabelle A 19 Verteilung des Item 10 des LIST-Inventars.....	x
Tabelle A 20 Deskriptive Statistik des summierten Index der LIST-Items	x
Tabelle A 21 Ergebnisse der Varianzanalyse der LIST-Items	xi
Tabelle A 22 Verteilung des Items Sicherheit zu den PES-Items nach Skalenlayout	xii
Tabelle A 23 Verteilung des Items Wichtigkeit des Themas der PES-Items nach Skalenlayout.....	xii
Tabelle A 24 Verteilung des Items Sicherheit zu den LIST-Items nach Skalenlayout	xiii
Tabelle A 25 Verteilung des Items Wichtigkeit des Themas der LIST-Items nach Skalenlayout.....	xiii
Tabelle A 26 Ergebnisse der Varianzanalyse der Items Sicherheit und Wichtigkeit zu den Themen der PES und des LIST-Inventars nach Skalenlayout	xiv
Tabelle A 27 Verteilung der dichotomen Variable der Einstellungsstärke zu den Inhalten der PES und des LIST-Inventars nach Skalenlayout	xiv
Tabelle A 28 Ergebnisse der Varianzanalyse der EU-Items	xv
Tabelle A 29 Verteilung des EU-Item 1	xvi
Tabelle A 30 Verteilung des EU-Item 2	xvi
Tabelle A 31 Verteilung des EU-Item 3	xvii
Tabelle A 32 Verteilung des EU-Item 4	xvii

Tabelle A 33 Verteilung des EU-Item 5	xviii
Tabelle A 34 Verteilung des EU-Item 6	xviii
Tabelle A 35 Verteilung des EU-Item 7	xix
Tabelle A 36 Verteilung des EU-Item 8	xix
Tabelle A 37 Verteilung des EU-Item 9	xx
Tabelle A 38 Verteilung des EU-Item 10	xx
Tabelle A 39 Deskriptive Statistik des summierten Index der EU-Items	xxi
Tabelle A 40 Ergebnisse der logistischen Multilevel-Regression zur Wahl (a) der rechten Extremkategorie, (b) der Mittelkategorie und (c) der DK-Kategorie mit Interaktion „Bildung×Layout“	xxii
Tabelle A 41 Ergebnisse der logistischen Multilevel-Regression zur Wahl (a) der rechten Extremkategorie, (b) der Mittelkategorie und (c) der DK-Kategorie mit Interaktion „Altersgruppe×Layout“	xxiii

I Einleitung

Antwortskalen sind die Messlatten der Umfrageforschung. Befragte Personen nutzen sie, um ihre Antworten zu den gestellten Fragen zu berichten. Dabei kommen unterschiedliche Typen von Antwortskalen zum Einsatz, wobei in der quantitativen Sozialforschung vor allem Ratingskalen sehr häufig Verwendung finden. Der Vorteil von Ratingskalen ist, dass sie ein zugrunde liegendes Bewertungskontinuum (z.B. Zustimmung, Intensität, Zufriedenheit) in wenige Bewertungen unterteilen, anhand derer die befragten Personen ihre Einstellung zu Objekten, Themen etc. kategorisieren können (Likert 1932; Thurstone und Chave 1929). Beispielsweise wird die Zustimmung zu einer Frage in die fünf Bewertungskategorien „stimme sehr zu – stimme zu – weder noch – lehne ab – lehne sehr ab“ unterteilt. Somit werden die Antworten durch die Ratingskala standardisiert erfasst, wodurch der Vergleich der gemessenen Einstellungen und Bewertungen sowie der Einsatz multivariater Methoden zu ihrer Analyse möglich sind (Franzen 2014; Schnell et al. 2011). Zu den fast unzähligen Gestaltungsaspekten von Ratingskalen, wie beispielsweise der Anzahl der Antwortkategorien, der Verwendung verbaler und / oder numerischer Labels oder der Skalenpolarität, gibt es eine reichhaltige Forschung (für einen Überblick vgl. u.a. Menold und Bogner 2012). Ziel dieser Forschungen ist, den Einfluss der verschiedenen Gestaltungsaspekte im Einzelnen oder in ihrem Zusammenspiel auf das Antwortverhalten von Befragten herauszuarbeiten und zu untersuchen, wie dadurch die Qualität der Antworten beeinflusst wird.

Seit einigen Jahren kommen Online-Befragungen in der quantitativen Sozialforschung mehr und mehr zum Einsatz. Da Online-Befragungen eine Vielzahl an Gestaltungsmöglichkeiten bieten, sind mit ihrer zunehmenden Popularität auch die Effekte der graphischen Gestaltung von Fragebogenelementen - auch als „Layouteffekte“ bezeichnet - in den Fokus der methodischen Forschung gerückt. Verschiedene Studien betrachten beispielsweise die graphische Gestaltung von Skip-Anweisungen (Befragte sollen Fragen überspringen), Datumsformaten oder

Antwortskalen (vgl. u.a. Christian et al. 2007; Smyth et al. 2006; Toepoel und Dillman 2011; Tourangeau et al. 2007). Für Ratingskalen ist - unter anderem - die Frage nach der Gestaltung der nicht-inhaltlichen Antwortkategorie (z.B. „weiß nicht“, „keine Meinung“, „kann ich nicht sagen“, kurz „DK-Kategorie“) und der damit einhergehenden Effekte auf das Antwortverhalten der befragten Personen offen. Dies ist der Forschungsgegenstand dieser Dissertation. Für andere gestalterischen Elemente von Ratingskalen ist bereits bekannt, dass sie die befragten Personen in ihrer Interpretation und Nutzung der Skala systematisch beeinflussen können (Smith 1995; Toepoel et al. 2008,2009; Toepoel und Dillman 2011).

Tourangeau et al. (2004) haben einen theoretischen Erklärungsansatz für das Auftreten von Layouteffekten in Ratingskalen entwickelt: Sie nehmen an, Befragte wenden bei Fragebögen, die sie eigenständig ohne Interviewer ausfüllen, interpretative Heuristiken an, wie beispielsweise „left and top means first“. Damit schreiben sie den nonverbalen, graphischen Ratingkalenelementen eine inhaltliche Bedeutung zu und nutzen diese Elemente im Fragebeantwortungsprozess als zusätzliche Informationsquelle zu den Fragen. Für die Darstellung der DK-Kategorie ist die interpretative Heuristik „middle means typical“ von besonderer Bedeutung. Diese Heuristik besagt, dass die optische, visuelle Mittelkategorie einer Skala als typischer, allgemeiner Wert betrachtet wird. Sie ist daher im Beantwortungsprozess ein Referenzpunkt für die Befragten, von dem aus sie ihre eigene Einstellung oder Bewertung herleiten. Wird die DK-Kategorie ohne eine graphische Abgrenzung rechts in die Skala integriert, also an das rechte Skalenende angehängt, wird der visuelle Skalenmittelpunkt verschoben und stimmt deshalb nicht mehr mit dem inhaltlichen Mittelpunkt der Skala überein. Wenn befragte Personen die Ratingskala eher visuell als verbal interpretieren, ziehen sie eine falsche Schlussfolgerung über die inhaltliche Skalenmitte, was in einer Verschiebung der Antworten, ebenfalls hin zum rechten Skalenende, resultieren kann. Diesen Effekt konnten Tourangeau et al. (2004) in einem Online-Experiment zeigen. Sie vergleichen dabei diese integrierte Darstellung mit einer, in der die DK-Kategorie mittels eines Trennstrichs graphisch von den inhaltlichen Antwortkategorien der Ratingskala separiert ist. In dieser abgetrennten Darstellung stimmen der visuelle und der inhaltliche Mittelpunkt der Skala miteinander überein, weshalb Befragte in ihrer Wahrnehmung und Interpretation der Skala nicht

fehlgeleitet werden sollten. Entsprechend der Annahmen beobachten Tourangeau et al. (2004) einen höheren Mittelwert in der angehängten Darstellung der DK-Kategorie, der durch eine Verschiebung der Antworten hin zum rechten Skalenende bedingt ist. Befragte scheinen durch die visuelle Gestaltung der Ratingskala in ihrem Antwortverhalten beeinflusst zu werden. Für die abgetrennte Darstellung zeigt sich ebenfalls ein Layouteffekt: die DK-Kategorie scheint durch ihre Abtrennung mittels eines Trennstrichs von den Befragten betont wahrgenommen zu werden, was zu einer höheren Rate an Item Nonresponse führt als in der angehängten Darstellung der DK-Kategorie. Diesen Effekt konnte die Studie von Christian et al. (2009) nicht bestätigen. Allerdings zeigt ein Vergleich beider Studiendesigns deutlich, dass erhebliche Unterschiede, unter anderem im Fragethema, in der Verwendung verbaler und numerischer Labels und in der untersuchten Bevölkerung vorhanden sind, die für die uneinheitlichen Ergebnisse hinsichtlich der Layouteffekte verantwortlich sein können (Christian et al. 2009).

Widersprüchliche Ergebnisse sind für die Praxis der empirischen Forschung problematisch, weil keine konkreten Gestaltungsempfehlungen (so genannte Best Practice) für das Layout von Ratingskalen und deren DK-Kategorie ausgesprochen werden können. Der Stand der Forschung fordert also zu vertiefter Untersuchung heraus. Man kann festhalten: Selbst wenn dieselbe Frage gestellt und inhaltlich dieselbe Ratingskala zur Antwortabgabe vorgegeben wurde, kann eine unterschiedliche graphische Gestaltung der DK-Kategorie der Ratingskala die Befragten systematisch beeinflussen. Dadurch kommt es zu einer Verminderung der Qualität der Messung und zu einer Gefährdung der Vergleichbarkeit der Daten. Darüber hinaus scheinen Layouteffekte durch weitere Faktoren des Befragungskontextes bedingt zu werden, wozu bisher kaum systematische Forschung betrieben wurde. Diese Dissertation setzt sich daher zum Ziel, weitere Kenntnisse über das Auftreten von Effekten, bedingt durch das Layout der DK-Kategorie in Ratingskalen, und über ihre Kovariate zu gewinnen. Es werden die Fragen verfolgt, ob

1. Layouteffekte gleichermaßen bei Items mit unterschiedlichen thematischen Inhalten (politische Einstellungen vs. Alltagsverhalten) auftreten,
2. Layouteffekte durch die Einstellungsstärke der Befragten gegenüber den Iteminhalten moderiert werden und

3. die beiden persönlichen Merkmale „Bildung“ und „Alter“ als Indikatoren für die kognitive Leistungsfähigkeit Kovariate von Layouteffekten sind.

Hierzu werden drei Ratingskalenlayouts untersucht: Im ersten Layout wird die DK-Kategorie mittels eines Trennstrichs von den inhaltlichen Kategorien abgetrennt, im zweiten Layout wird die DK-Kategorie an die Ratingskala angehängt und im dritten Layout wird keine DK-Kategorie vorgegeben. Als abhängige Variable werden nicht wie in den beiden existierenden Studien (Christian et al. 2009; Tourangeau et al. 2004) Antwortverteilungen einzelner Items betrachtet. In dieser Untersuchung wird der Einfluss des Ratingskalenlayouts auf das Antwortverhalten über mehrere Items, die verschiedene latente Konstrukte repräsentieren, hinweg betrachtet. Basierend auf den Forschungshypothesen wird das Antwortverhalten als „Wahl der rechten Extremkategorie“, „Wahl der Mittelkategorie“ und „Wahl der DK-Kategorie“ operationalisiert. Bei diesem Vorgehen kann auch der Einfluss der potentiellen Kovariaten „Einstellungsstärke“, „Bildung“ und „Alter“ auf das Auftreten und auf die Stärke der Layouteffekte untersucht werden. Durch die Einbeziehung einer Ratingskala ohne DK-Kategorie wird es zudem möglich, Aussagen über den Einfluss einer nicht vorgegebenen DK-Kategorie auf das Antwortverhalten der Befragten zu machen. Somit stellt diese Dissertation in manchen Teilen eine Replikation der bestehenden Studien dar. Gleichzeitig bildet sie aber auch eine Erweiterung dieser Studien, da sie durch die Untersuchung potentieller Kovariate von Layouteffekten versucht, einen Erklärungsbeitrag dafür zu leisten, wieso der aktuelle Forschungsstand widersprüchliche Ergebnisse umfasst. Zudem handelt es sich bei dieser Untersuchung von Layouteffekten der DK-Kategorie um die erste im deutschsprachigen Raum.

Diese Dissertation basiert auf den Daten zweier Experimente. Teilnehmer des ersten Experiments waren Studenten zweier deutscher Universitäten, deren Einstellungen zu verschiedenen politischen und persönlichen Themen mittels Selbstausfüller-Fragebögen erhoben wurden. Das zweite Experiment war Teil einer Online-Befragung des GESIS Online Access Panel Pilots, welche auf einer heterogenen, deutschsprachigen Erwachsenenstichprobe basiert. Die Auswertung der Daten erfolgt mittels des multivariaten Verfahrens der logistischen Multilevel-Regression mit fixen Effekten. Die Ergebnisinterpretation wird anhand durchschnittlicher marginaler Effekte vorgenommen. Somit ist es möglich, Aussagen darüber zu -

treffen, wie wahrscheinliche Befragte bei einem bestimmten Ratingskalenlayout die drei untersuchten Antwortkategorien wählen und ob diese Unterschiede signifikant sind.

Die Dissertation ist in fünf Kapitel unterteilt: Das erste Kapitel umfasst diese Einleitung.

Im zweiten Kapitel wird auf theoretischer Basis geschildert, warum Befragte durch das Layout vor allem in Erhebungsmodi, in denen sie selbständig ohne Interviewer das Erhebungsinstrument ausfüllen, in ihrer Antwortabgabe beeinflusst werden. Hierbei wird auf den eingeschränkten konversationalen Kontext selbstadministrierter Befragungen eingegangen und daraus abgeleitet, warum Befragte die unterschiedlichen Elemente eines Fragebogens als Information bewerten und in den Fragebeantwortungsprozess miteinbeziehen. Ebenfalls in diesem Kapitel werden die interpretativen Heuristiken nach Tourangeau et al. (2004) als theoretischer Erklärungsansatz für das Auftreten von Layouteffekten und der Stand der Forschung zu Layouteffekten und ihrer Kovariate ausführlich vorgestellt. Des Weiteren werden die theoretischen Argumente für und wider der Vorgabe von Don't Know-Kategorien in Ratingskalen diskutiert und der aktuelle Forschungsstand dazu wiedergegeben. Die Effekte des Layouts der DK-Kategorie in Ratingskalen werden anhand der beiden Studien von Tourangeau et al. (2004) und Christian et al. (2009) beschrieben. Hierbei erfolgt auch eine systematische Gegenüberstellung beider Studien mit dem Ziel, potentielle Kovariate von Layouteffekten zu identifizieren. Schließlich werden die Forschungsziele und Forschungshypothesen der Studie 1 und der Studie 2 dieser Dissertation vorgestellt.

Das dritte Kapitel der Dissertation umfasst die empirische Studie 1, die sich mit dem Einfluss des Layouts auf das Antwortverhalten der Befragten und der moderierenden Wirkung der individuellen Einstellungsstärke auf das Auftreten der Layouteffekte auseinandersetzt. Studie 1 basiert auf den Daten der papierbasierten Befragung von Studenten der Universität Mannheim und der Universität Gießen, deren Erhebung und Statistik im Rahmen der Vorstellung des Studiendesigns erläutert wird. Hierbei wird auch eine ausführliche Darstellung der verwendeten Methode, der logistischen Multilevel-Regression mit fixen Effekten sowie der Ergebnisinterpretation anhand durchschnittlicher marginaler Effekte gegeben. Im

Anschluss werden die deskriptiven, bivariaten und multivariaten Ergebnisse vorgestellt und diskutiert.

Das vierte Kapitel behandelt die empirische Studie 2, welche untersucht, ob die in Studie 1 erzielten Ergebnisse anhand einer heterogenen Erwachsenenpopulation im Web-Modus repliziert werden können. Auch ist von Interesse, welchen moderierenden Einfluss die persönlichen Merkmale „Bildung“ und „Alter“ auf die Empfänglichkeit für das Ratingskalenlayout nehmen. Das vierte Kapitel ist strukturell wie das dritte Kapitel aufgebaut: Zunächst erfolgt eine Beschreibung des Studiendesigns. Im Anschluss daran werden die deskriptiven, bivariaten und multivariaten Ergebnisse präsentiert und diskutiert.

Im fünften und letzten Kapitel wird eine zusammenfassende Schlussbetrachtung der Dissertation vorgenommen. Dabei werden die zentralen Ergebnisse im Hinblick auf die Forschungsziele und -hypothesen zusammengefasst und diskutiert. Aus den zentralen Ergebnissen abgeleitet ergeben sich Empfehlungen für künftige Forschung zu Layouteffekten sowie Layoutempfehlungen für die empirische Forschungspraxis.

II Theorie & Stand der Forschung

1. Bedeutung des Layouts von Ratingskalen

Dieser erste Abschnitt des zweiten Kapitels befasst sich mit dem Layout von Ratingskalen, also ihrer visuellen Gestalt, die sich durch die Anordnung graphischer Elemente, wie beispielsweise Linien, Schattierungen und Symbole der Antwortkategorien, zueinander ergibt. Die graphischen Elemente werden dabei der Gruppe der nonverbalen Skalenelemente zugeordnet und von den verbalen Skalenelementen unterschieden. Diese Klassifikation ist Thema des ersten Unterabschnitts (1.1). Im zweiten Unterabschnitt (1.2) wird der Befragungsprozess als spezielle Form der Konversation betrachtet, dem, wie auch einer alltäglichen Konversation, Prinzipien zugrunde liegen, deren Einhaltung von den Beteiligten erwartet wird. Daraus folgt, dass Befragte, vor allem in selbstadministrierten Umfragen, auch nonverbale, gestalterische Elemente von Ratingskalen als Information werten und sich von diesen in ihrem Antwortverhalten beeinflussen lassen können. Schließlich wird im dritten Unterabschnitt (1.3) mit den interpretativen Heuristiken von Tourangeau et al. (2004) ein theoretischer Ansatz vorgestellt, der eine Erklärung dazu anführt, wie Befragte nonverbale Elemente von Fragebögen auffassen und welche Folgen das auf ihre Interpretation der Ratingskala nehmen kann.

1.1 Verbale und nonverbale Ratingskalenelemente

Ratingskalen sind aus verschiedenen Elementen aufgebaut. Üblicherweise bestehen sie aus einer bestimmten Anzahl von Antwortkategorien - in der Regel vier bis elf - die mit numerischen und / oder verbalen Etiketten versehen sein können. Die visuelle Gestalt der Skala, also ihr Layout, ist aus einer Vielzahl von graphischen Elementen und deren graphischen Anordnung zueinander aufgebaut. Zu diesen

Elementen in Ratingskalen zählen unter anderem Antwortkästchen oder Radio-Buttons, Linien, einfarbige oder bunte Schattierungen oder Schriftarten und Schriftgrößen.

Die unterschiedlichen Elemente von Ratingskalen, die auch in Fragebögen generell Verwendung finden, werden in der sozialwissenschaftlichen Umfrageforschung in verbale und nonverbale Elemente unterteilt. Verbale Elemente sind gesprochene oder geschriebene Worte. Nonverbale Elemente umfassen dann alle nicht verbalen Elemente und werden weiter unterschieden in numerische, graphische und symbolische Elemente (Christian et al. 2009; Jenkins und Dillman 1995). Numerische Elemente werden hauptsächlich für die numerischen Etiketten der Antwortskalen eingesetzt, also als Messinstrument. Symbolische Elemente umfassen kulturell bedeutsame Zeichen, die zu Kommunikationszwecken eingesetzt werden, wie beispielsweise ein Pfeil, der zur nächsten Frage weiterleitet. Graphische Elemente von Fragebögen beinhalten schließlich Elemente wie die Schriftgröße, Helligkeit und Farbe, Form, Lage oder die räumliche Anordnung von Wörtern, Zahlen und Symbolen (Christian und Dillman 2004).

Couper et al. (2004a) differenzieren die verbalen und nonverbalen Elemente von Fragebögen noch dahingehend, ob sie wesentlich zur Fragebeantwortung nötig sind - sogenannte Aufgabenelemente - oder ob sie lediglich zu Zwecken der Fragebogengestaltung eingesetzt werden - sogenannte Gestaltungselemente. Aufgabenelemente sind die Frageformulierungen, verbale Etiketten in Antwortskalen und Instruktionen. In den meisten Fällen entstammen sie der Gruppe der verbalen Elemente. Allerdings können auch nonverbale Elemente dieser Gruppe zugehören, wie zum Beispiel Bilder von Portionsgrößen in Diätumfragen. Im Gegensatz dazu sind Gestaltungselemente all jene Elemente, die nicht wesentlich für den Beantwortungsprozess sind, welche dem Fragebogen aber sein Design geben. Meistens sind Gestaltungselemente nonverbale, bildliche Merkmale, wie zum Beispiel das Studienlogo, Hintergrundfarbe oder das Layout von Antwortskalen. Es wird angenommen, dass Gestaltungselemente zu einem Design des Fragebogens beitragen, welches die Befragten zu einer Teilnahme motiviert (Couper et al. 2004a).

1.2 Ratingskalenelemente im konversationalen Kontext von Umfragen

Die obige Darstellung verdeutlicht, dass Ratingskalen neben verbalen auch aus einer Vielzahl unterschiedlicher nonverbaler Elemente bestehen. Welche Bedeutung diesen nonverbalen Elementen im Fragebeantwortungsprozess zugesprochen werden muss und welchen Einfluss sie auf das Antwortverhalten der Befragten nehmen, hängt maßgeblich vom Modus der Datenerhebung ab. Denn wird in Face-to-Face- als auch in telefonischen Interviews vorwiegend der Audiokanal und somit verbale Sprache genutzt, um dem Befragten die nötige Information zu übergeben und mit ihm zu kommunizieren, liegt Befragten in selbstadministrierten Umfragen nahezu ausschließlich visuell präsentierte Information vor. Nur der visuelle Kommunikationskanal dient in selbstadministrierten Umfragen als Informationsübertragung zwischen Forscher und Befragten. Diese nehmen neben der verbalen, geschriebenen Sprache auch visuell die nonverbalen Fragebogenelemente wahr (Redline und Dillman 2002).

Somit unterscheiden sich auch die konversationalen Kontexte von selbstadministrierten und Interviewer geführten Umfragen maßgeblich voneinander: Während in Face-to-Face- oder telefonischen Interviews der Interviewer als Kommunikator sowohl die Fragen als auch die Intention und Erwartung des Forschers übermittelt, stellt in selbstadministrierten Umfragen das Befragungsinstrument selbst das einzige Mittel der Kommunikation zwischen Forscher und Befragtem dar (Christian und Dillman 2004; Couper et al. 2004a). Befragungsinstrumente selbstadministrierter Umfragen sind meist Papierfragebögen oder Fragebögen, die auf dem Computer wiedergegeben werden (computerisierte Fragebögen). In dieser Befragungssituation können Befragte bei Problemen mit dem Befragungsinstrument, wie beispielsweise zum Frageverständnis, der Benutzung der vorgegeben Antwortskala oder anderen Unklarheiten, nicht auf zusätzliche, unterstützende Informationen von Seiten des Interviewers zurückgreifen. Eine selbstadministrierte Umfrage stellt also einen stark eingeschränkten konversationalen Kontext dar. Trotzdem wird diese Situation von Befragten als eine Konversation definiert und dementsprechend erwarten sie die Einhaltung des kooperativen Prinzips, welches Grice (1993) in vier Maximen der Konversation wie folgt beschreibt:

1. *Quantität*

Von jedem Beitrag zu einer Konversation wird erwartet, dass er mindestens so informativ ist, wie es für den anerkannten Zweck des Gesprächs nötig ist, aber gleichzeitig sollte der Beitrag wiederum nicht informativer als nötig sein.

2. *Qualität*

Der Gesprächsbeitrag sollte nichts beinhalten, was vom Redner als falsch angenommen wird oder für das er nicht genügend Anhaltspunkte besitzt.

3. *Relevanz*

Von einem Gesprächsbeitrag wird erwartet, dass er für das Thema des Gesprächs relevant ist.

4. *Modalität*

Es wird von einem Beitrag zu einer Konversation erwartet, dass er klar, eindeutig und geordnet erfolgt.

Im Konversationsmodell von Grice (1993) wird die Annahme verfolgt, dass es eine allgemeingültige, akzeptierte Art des Sprechens gibt, welche als Standardverhalten in einer Konversation erwartet wird. Hören wir eine Äußerung, erwarten wir, dass diese grundsätzlich informativ, wahrheitsgemäß, relevant und eindeutig ist. Scheint diese Äußerung nicht diesem Modell zu entsprechen, nehmen wir nicht an, dass die Äußerung unsinnig ist, sondern dass aus ihr auf einen bestimmten Sinn geschlossen werden kann (Davies 2000).

Bless et al. (1993) übertragen diese Maxime der Konversation auf den Kontext von standardisierten Umfragen. Sie zeigen in Experimenten, dass Befragte in einer stark eingeschränkten konversationalen Situation, wie der selbstadministrierten Befragung, alle vom Forscher angebotene Information als aufgabenrelevant betrachten und sich bemühen, ihnen eine Funktion und einen Sinn zuzuschreiben. Befragte entnehmen den Elementen des Befragungsinstruments jede erhältliche Information und betrachten sie als relevant zur Erfüllung der Aufgabe. Das sind sowohl verbale als auch nonverbale Elemente des Befragungsinstruments. Auch wenn der Fragebogenentwickler Elemente wie Farbe, Schattierungen oder die graphische Anordnung der Antwortkategorien lediglich zu Gestaltungszwecken einsetzt, können Befragte annehmen, dass diesen Elementen eine Bedeutung

zukommt und sie dann im Befragungsprozess miteinbeziehen (Bless et al. 1993; Couper et al. 2004a). Besonders wenn Befragte unsicher hinsichtlich der Fragestellung oder Fragebeantwortung sind, greifen sie auf Informationen und Hinweise aus dem Kontext der Frage und Antwortskala zurück, wie beispielsweise numerische Etiketten, die Abstände zwischen den einzelnen Skalenpunkten oder die Gestaltungsform der Antwortskala (Schwarz et al. 1991; Smith 1995). Tourangeau et al. (2007) bezeichnen diesen Rückgriff auf nonverbale Kontextinformationen als visuelle Kontexteffekte. Daher stellen in selbstadministrierten Umfragen, anders als in Interviews, neben Worten zusätzlich auch die nonverbalen Elemente eine wichtige Quelle zur Erschließung der Fragebedeutung dar (Christian und Dillman 2004).

1.3 Heuristiken der Interpretation nonverbaler Ratingskalenelemente

Einen ersten systematischen Erklärungsansatz der Effekte nonverbaler Fragebogenelemente stellen Tourangeau et al. (2004, 2007) vor. Sie nehmen dabei Bezug auf Erkenntnisse der visuellen Wahrnehmungsforschung, wonach visuelle Wahrnehmung kein passiver, sondern vielmehr ein intelligenter Prozess aktiver Konstruktion ist. Eine zentrale Rolle in diesem unbewusst stattfindenden, visuellen Konstruktionsprozess spielen grundlegende Fähigkeiten wie Tiefen- oder Formwahrnehmung und interpretative Heuristiken, die angewandt werden, um nonverbalen Elementen einer Frage oder einer Antwortskala eine Bedeutung zuzuschreiben. Die Autoren führen konkret fünf interpretative Heuristiken auf, die Befragte bei der Interpretation der nonverbalen Elemente in Fragebögen heranziehen:

1. Middle means typical

Die mittlere Antwortalternative in einer Liste bzw. die Mittelkategorie in einer Skala wird von Befragten als der typische Wert angenommen. Weil die mittlere Option als typisch betrachtet wird, stellt sie für Befragte im Beantwortungsprozess einen Referenzpunkt dar, von dem aus sie ihre eigene Position beurteilen.

2. Left and top means first

Die linke bzw. oberste Antwortalternative repräsentiert die erste Alternative aus inhaltlicher Sicht. Diese Heuristik hat ihren Ursprung in der Lesereihenfolge der

meisten Sprachen westlicher Länder. Für Antwortkategorien von Skalen wird daher erwartet, dass einer der beiden Endpole links bzw. oben in der Reihenfolge steht, der entgegengesetzte Pol dann wiederum rechts bzw. unten, und die Kategorien dazwischen sollten einer logischen Abfolge unterliegen. Stimmt die Anordnung nicht mit dieser Annahme überein, verwirrt dies die Befragten und sie machen eher Fehler.

3. *Near means related*

Sind Items oder Antwortkategorien räumlich nah beieinander dargestellt, zum Beispiel Items auf einer Bildschirmseite oder benachbarte Antwortkategorien, schließen Befragte auf eine inhaltliche Nähe. Diese Heuristik basiert auf dem Prinzip der Nähe der Gestaltpsychologie nach Koffka (1935) und Wertheimer (1923). Nach diesem Prinzip tendieren Menschen dazu, nah beieinander liegende Objekte als Einheiten zusammengefasst wahrzunehmen.

4. *Up means good*

„Up means good“ stellt eine Variante der zweiten Heuristik „left and top means first“ dar und besagt, dass bei einer vertikalen Liste die oberste Alternative mit der wünschenswerten Alternative gleichgesetzt wird. Sie geht auf kulturell tief verwurzelte Metaphern, wie „Himmel=oben“ / „Hölle=unten“, „Aufstieg = Erfolg“, „unten sein“ zurück. Angewendet auf Ratingskalen ist daraus zu schließen, dass Befragte die zustimmende Antwortkategorie an oberster Stelle erwarten. Die Annahme ist, dass diese Erwartung Befragte bei der Beantwortung nicht völlig fehlleiten kann, aber dass der Beantwortungsprozess einfacher für den Befragten ist, wenn sie erfüllt wird.

5. *Like (in appearance) means close (in meaning)*

Antwortoptionen, die graphisch gleich oder ähnlich erscheinen, hängen auch inhaltlich zusammen. Diese Heuristik geht zurück auf das Prinzip der Ähnlichkeit der Gestaltpsychologie von Koffka (1935) und Wertheimer (1923), wonach ähnliche Objekte zu Einheiten gruppiert werden.

Tourangeau et al. (2004, 2007) haben diese Heuristiken in Web Survey-Experimenten empirisch untersucht und kommen zu dem Ergebnis, dass ähnlich dem Vorgang einer optischen Täuschung, auch die Befolgung der Heuristiken der Interpretation nonverbaler Fragebogenelemente zu einer systematischen

Fehlauffassung der Frage oder Antwortskala, und somit zu systematischen Messfehlern, führen kann. Nonverbale Skalenelemente, die lediglich zu Gestaltungszwecken eingesetzt wurden, können somit von der wichtigeren, verbalen Information ablenken oder gar die Basis für ungewollte Schlussfolgerungen darstellen, welche zu einer Fehlinterpretation der verbalen Information führt (Couper et al. 2004a).

2. Effekte des Layouts von Ratingskalen

Die reiche Palette an Gestaltungsmöglichkeiten in Web Surveys macht es notwendig, bessere Kenntnisse über die Effekte nonverbaler Fragebogenelemente zu gewinnen. Dieser Notwendigkeit trägt eine stetig wachsende Anzahl an Forschungsstudien Rechnung. Während über die Verwendung nonverbaler, numerischer Information (hauptsächlich zu numerischen Etiketten in Ratingskalen) eine umfangreiche Forschung existiert (vgl. u.a. Amoo und Friedman 2001; Armitage und Deeproose 2004; Hartley und Betts 2010; Krosnick und Fabrigar 1997; Schwarz et al. 1991), ist die Studienlage zu den Effekten der graphischen Gestaltung von Ratingskalen überschaubar und wird im folgenden Unterabschnitt (2.1) vorgestellt. Der zweite Unterabschnitt (2.2) bietet einen Überblick über die empirischen Befunde über die Kovariate von Layouteffekten, die entweder in den Eigenschaften der Befragten oder in den Eigenschaften der Ratingskala begründet sind. Die beiden, für diese Dissertation zentralen Publikationen über Untersuchungsergebnisse zu Effekten der graphischen Darstellung der DK-Kategorie in Ratingskalen, werden gesondert in Abschnitt (4) ausführlich vorgestellt und diskutiert.

2.1 Stand der Forschung

Die graphische Anordnung der Antwortkategorien einer Ratingskala kann unterschiedlich in Fragebögen umgesetzt werden. Am häufigsten werden Ratingskalen als horizontal oder vertikal ausgerichtete, lineare Anordnung der Kategorien dargestellt. Jedoch existieren auch weitere Varianten, wie beispielsweise

eine Darstellung in mehreren Spalten oder in mehreren Zeilen. Die empirischen Befunde über die Effekte der graphischen Anordnung der Antwortkategorien sind gemischt. Für den Vergleich von linearen versus nicht-linearen, mehrspaltigen Darstellungsformen (sowohl in selbstadministrierten Papierfragebögen als auch in Web Surveys) werden signifikant unterschiedliche Antwortverteilungen und Mittelwerte nachgewiesen (Christian und Dillman 2004; Christian et al. 2009; Toepoel et al. 2009). Sind die Antwortkategorien nicht-linear, mehrzeilig anstatt linear, vertikal dargestellt, wählen Befragte häufiger Kategorien aus der oberen Reihe (Christian und Dillman 2004; Christian et al. 2009). Auch sind für diese Darstellungsform längere Antwortzeiten zu beobachten, was auf eine erschwerte mentale Verarbeitung hindeutet (Christian et al. 2009). Weder die Tendenz zur Wahl der oberen Kategorien noch die längeren Antwortzeiten zu den nicht-linearen, mehrzeiligen Skalenlayouts werden hingegen in der Studie von Toepoel et al. (2009) nachgewiesen.

Für den Vergleich von vertikal versus horizontal dargestellten Ratingskalen werden signifikant unterschiedliche Antwortverteilungen beobachtet (Friedman und Friedman 1994; Toepoel et al. 2009). Allerdings kann keine eindeutige Effektrichtung ausgemacht werden. Verschiedene Studien belegen aber, dass in vertikal ausgerichteten Ratingskalen vermehrt Primacy-Effekte auftreten und daher eher eine horizontale Darstellung genutzt werden sollte (Tourangeau et al. 2000).

Die Skalenorientierung und die Kombination von verbalen und numerischen Etiketten beeinflussen ebenfalls das Antwortverhalten der Befragten. Hartley und Betts (2010) zeigen in einem Web Survey-Experiment, dass höhere Ratings erzielt werden, wenn eine Ratingskala mit dem zustimmenden verbalen Etikett und dem höchsten numerischen Wert beginnt. Als Vergleich dient eine Skala, die mit dem ablehnenden verbalen Etikett und dem niedrigsten numerischen Wert beginnt. Dieses Ergebnis steht in Einklang mit der Heuristik „up means good“, gemäß der Befragte erwarten, dass die erste Antwortkategorie die zustimmende ist (Tourangeau et al. 2004).

Den Effekt einer umgekehrten Reihenfolge der Antwortkategorien („täglich“ bis „einmal pro Woche“ versus „einmal pro Woche“ bis „täglich“) untersuchen Stern et al. (2007) in einer postalischen Umfrage. Für eines der beiden untersuchten Items sind signifikante Unterschiede in den Antwortverteilungen zu beobachten; die

beiden äußersten Kategorien „täglich“ bzw. „einmal pro Woche“ werden häufiger gewählt, wenn sie am Skalenanfang anstatt am Skalenende dargestellt sind. Dieses Ergebnis steht im Einklang mit der Literatur zu Primacy-Effekten (vgl. u.a. Bishop und Smith 2001; Galesic et al. 2008b; Krosnick und Alwin 1987).

Verschiedene Studien zeigen, dass sich Befragte bei der Interpretation der Skala am visuellen Skalenmittelpunkt orientieren. Kommt es zu einer Diskrepanz zwischen dem inhaltlichen und dem visuellen Skalenmittelpunkt (weil z.B. ungleichmäßige Abstände zwischen den Antwortkategorien herrschen), resultiert dies in signifikant unterschiedlichen Antwortverteilungen. Dieses Ergebnis steht im Einklang mit den Annahmen der Heuristik „middle means typical“ (Christian et al. 2009; Tourangeau et al. 2004).

Hinsichtlich der farblichen Gestaltung von Ratingskalen zeigt sich, dass bei Verwendung von unterschiedlichen farblichen Schattierungen an den beiden Skalenenden (z.B. Blautöne auf der ablehnenden Seite, Rottöne auf der zustimmenden Seite), Befragte den ablehnenden Bereich eher meiden als wenn an beiden Enden Schattierungen desselben Farbtons eingesetzt werden. Dieser Effekt trat nicht mehr auf, wenn die Skala mit verbalen Etiketten versehen wurde (Toepoel und Dillman 2011; Tourangeau et al. 2007). Dieses Ergebnis steht im Einklang mit der Heuristik „like (in appearance) means close (in meaning)“, gemäß der Befragte die Antwortkategorien als inhaltlich weiter voneinander entfernt wahrnehmen, wenn diese sich in ihrer Gestaltung voneinander unterscheiden (Tourangeau et al. 2004). Eine ähnliche Untersuchung ist aus der Forschung zu numerischen Etiketten bekannt. Hier wird beobachtet, dass Befragte die Kategorien der unteren Hälfte einer Ratingskala meiden, wenn diese mit negativen Zahlen versehen sind (Schwarz et al. 1991).

2.2 Kovariate von Layouteffekten

Die Ergebnisse der Studien über die Effekte der verschiedenen nonverbalen, graphischen Elemente von Ratingskalen weisen darauf hin, dass Befragte mal mehr, mal weniger in ihrem Antwortverhalten von nonverbalen Elementen beeinflusst werden. Um erklären zu können, unter welchen Bedingungen Befragte durch nonverbale Elemente beeinflusst werden, setzten sich auch verschiedene Studien

mit Kovariaten von Layouteffekten auseinander, deren Ergebnisse im Folgenden vorgestellt werden. Hierbei erfolgt eine Konzentration auf Kovariate, die in den Eigenschaften der Befragten (Unterabschnitt 2.2.1) oder in den Eigenschaften der Ratingskala (Unterabschnitt 2.2.2) begründet sind.

2.2.1 Eigenschaften der Befragten

Demographische Charakteristika

Aus der Kognitionsforschung ist bekannt, dass die kognitive Leistungsfähigkeit mit zunehmendem Alter abnimmt (Salthouse 1991,2004). Daher wird angenommen, dass bei älteren im Vergleich zu jüngeren Befragten mehr Probleme mit dem Frageverständnis auftreten und diese daher eher auf weitere Informationen aus dem Umfeld der Frage zurückgreifen, um das Frageverständnis zu erleichtern. Dies können beispielsweise Informationen aus Inhalten vorhergehender Fragen, aus verbalen bzw. numerischen Etiketten oder anderen nonverbalen Elementen der Antwortskala sein (Schwarz et al. 1991). Stern et al. (2007) finden für Befragte über und unter 65 Jahren keine unterschiedlichen Layouteffekte. Toepoel et al. (2009) beobachten für Befragte über 65 Jahren stärkere Recency-Effekte, wenn eine horizontale mit einer vertikalen Ratingskala verglichen wird. Eine höhere Tendenz zu Response-Order-Effekten bei älteren Befragten ist aus der einschlägigen Literatur bekannt (vgl. u.a. Knäuper 1999).

In der Regel zieht eine geringere kognitive Leistungsfähigkeit einen niedrigeren formalen Bildungsabschluss nach sich (Asendorpf und Neyer 2012). Daher wird für Befragte mit einem niedrigen Bildungsabschluss erwartet, dass sie, aufgrund der geringeren kognitiven Leistungsfähigkeit, einer größeren Anstrengung bei der Bewertung der Antwortkategorien oder bei dem Frageverständnis unterliegen und daher eher ein Satisficing-Verhalten zeigen als Befragte mit einem hohen Bildungsabschluss. Gleichzeitig sind Befragte, die ein Satisficing-Verhalten zeigen empfänglicher für Effekte der nonverbalen Elemente im Fragekontext (Krosnick 1992). Deshalb wird angenommen, dass für Befragte mit einem niedrigen Bildungsabschluss eher Effekte des Skalenlayouts zu beobachten sind als für Befragte mit einem hohen Bildungsabschluss. Diese Annahme wird allerdings nicht in der Studie von Stern et al. (2007) unterstützt: für niedriger und höher Gebildete zeigen sich die gleichen Effekte des Skalenlayouts.

Aus den Erkenntnissen, dass Frauen häufiger die DK-Kategorie wählen, weil sie aufgrund ihrer Sozialisation eine stärkere Unsicherheit dabei verspüren, ihre Meinung öffentlich zum Ausdruck zu bringen als Männer (vgl. u.a. Rapoport 1982,1985), treffen Stern et al. (2007) die Schlussfolgerung, dass Frauen aus dieser Unsicherheit heraus eher auf nonverbale Elemente zurückgreifen als dies bei Männern der Fall ist, was sich empirisch aber ebenfalls nicht belegen lässt.

Kognitionsbedürfnis

Menschen unterscheiden sich in ihrem Bedürfnis nach Kognition; also danach, in welchem Ausmaß sie anstrengende kognitive Tätigkeiten betreiben und genießen. Diese Persönlichkeitseigenschaft wird in der Sozialpsychologie als „Need for Cognition“ bezeichnet und wurde 1982 von Cacioppo und Petty (1982) eingeführt. Gemäß dieses Modells bilden sich Menschen mit hohem Kognitionsbedürfnis ihre eigene Meinung durch das Abwägen all der ihnen vorliegenden Argumente. Menschen mit niedrigem Kognitionsbedürfnis hingegen lassen sich oft durch begleitende Signale beeinflussen und sind in ihrem Vorgehen nicht gleichsam gründlich (Aronson et al. 2008). Die Annahme ist, dass sich Befragte in ihrer Strategie bei der Informationssuche und Antwortabgabe im Fragebeantwortungsprozess nach ihrem Kognitionsbedürfnis unterscheiden und damit auch in ihrer Sensibilität für nonverbale Elemente. Befragte mit hohem Kognitionsbedürfnis tendieren eher dazu, alle ihnen zur Verfügung stehenden Informationen zu bewerten und für ihre Entscheidung zu Grunde zu legen als Befragte mit geringem Kognitionsbedürfnis. Deshalb sind sie auch sensibler für nonverbale Fragebogenelemente als Informationsquelle und nutzen diese eher bei der Fragebeantwortung, was auch empirisch belegt werden konnte (Toepoel et al. 2006).

Umfrage-Erfahrung

Es wird angenommen, dass Befragte, die bereits einige Erfahrung mit Umfragen besitzen, routiniert und sicher im Fragebeantwortungsprozess sind. Dadurch, so die Annahme, reagieren sie unsensibler auf nonverbale Elemente von Skalen als unerfahrene Personen (Toepoel et al. 2008). Ein Vergleich erfahrener mit neu rekrutierten Panelisten zeigt allerdings, dass beide Gruppen gleichermaßen auf die Variationen verschiedener Gestaltungsaspekte des Fragebogens, wie beispielsweise die Anzahl der Items pro Bildschirmseite oder die Reihenfolge der

Antwortkategorien, reagieren. Allerdings sind für die erfahrenen Panelisten schnellere Antwortzeiten, höhere Inter-Item-Korrelationen in Itembatterien und eine Tendenz zur Wahl der ersten Antwortkategorie zu beobachten, was den Schluss zulässt, dass diese Befragtengruppe eher zu Satisficing neigt als unerfahrene Panelisten (Toepoel et al. 2008).

Einstellungsstärke

Eine Einstellung ist eine „psychische Tendenz, die dadurch zum Ausdruck kommt, dass man ein bestimmtes Objekt mit einem gewissen Grad an Zuneigung oder Ablehnung bewertet“ (Eagly und Chaiken 1993, S.1). Einstellungen können durch zwei Maße bewertet werden: erstens durch ihre Valenz, also die Wertigkeit oder die Richtung, und zweitens durch ihre Stärke. Viele sozialwissenschaftliche Umfragen zielen darauf ab, die Valenz von Einstellungen zu messen, also ob ein Befragter eher positiv, neutral oder negativ gegenüber einem Objekt eingestellt ist. Die Messung der Stärke einer Einstellung hingegen ist oftmals nicht Ziel von Umfragen. Allerdings beeinflusst die Stärke einer Einstellung, wie empfänglich Befragte bei ihrer Antwortabgabe für den Kontext der Frage sind (Rockwood et al. 1997; Schwarz 1996; Tourangeau et al. 2004,2007).

Starke Einstellungen sind stabil und liegen klar auskristallisiert im Langzeitgedächtnis vor. Wird eine Person zu einem Thema befragt, zu dem sie eine starke Einstellung besitzt, ist der kognitive Prozess der Fragebeantwortung geradlinig: Die abgefragte Einstellung muss nicht erst anhand verschiedener, im Gedächtnis zur Verfügung stehender Informationen oder unter Zuhilfenahme von Informationen aus dem Kontext der Befragung generiert werden. Vielmehr kann die Einstellung direkt und ohne weitere Unterstützung aus dem Langzeitgedächtnis abgerufen werden. Somit ist für den Befragten einfach, die ihm klar vorliegende Einstellung anhand des vorgegebenen Antwortformats zu berichten (Bassili und Krosnick 2000). Anders verläuft der kognitive Prozess der Fragebeantwortung bei Themen, zu denen der Befragte lediglich eine schwache Einstellung besitzt. Schwache Einstellungen sind instabil und liegen nicht auskristallisiert im Langzeitgedächtnis vor. Der Befragte muss eine solche Einstellung situativ generieren, um sie berichten zu können. Dazu muss er zuerst eine relevante Beurteilung erbringen, was unter Zuhilfenahme von Informationen, die ihm im Gedächtnis vorliegen, geschieht. Oftmals werden dabei aber auch temporär

vorliegende Informationen aus dem Befragungskontext zu Hilfe genommen, welche die Befragten aus verbalen und nonverbalen Elementen der Frage und der Antwortskala gewinnen (Krosnick und Abelson 1991; Krosnick und Petty 1995; Schuman und Presser 1981; Schwarz et al. 1991).

In der empirischen Sozialforschung wird untersucht, ob Einstellungsstärke eine moderierende Wirkung auf die Empfänglichkeit der Befragten für Elemente aus dem Kontext besitzt. Dabei wird Einstellungsstärke meist anhand einzelner oder einiger weniger Attribute wie Extremität, Sicherheit, Intensität oder Wichtigkeit, die als erklärende Variable in die Analysen einbezogen werden, operationalisiert. Dabei zeigt sich, dass unsichere Befragte empfänglicher für verbale und nonverbale Elemente des Fragekontextes sind als solche, die sich sicher fühlen (Schwarz 1996; Tourangeau et al. 2004,2007). Rockwood et al. (1997) finden, dass Befragte bei ihnen eher unwichtigen Themen stärker auf verschiedene Elemente von Ratingskalen reagieren als bei ihnen sehr wichtigen Themen. Menold (2014) bezieht die Attribute „Sicherheit“ und „Wichtigkeit“ gegenüber der gemessenen Einstellung in ihre Studie über die Effekte verbaler Etiketten und der Anzahl von Antwortkategorien auf die Reliabilität von Skalen mit ein. Sie findet, dass mit zunehmender Einstellungsstärke, operationalisiert als multiplikativen Index der Attribute „Sicherheit“ und Wichtigkeit, die Tendenz zur Mitte abnimmt und die Tendenz zur Nutzung der extremeren Skalenkategorien zunimmt, wodurch die Reliabilität der Skalen ansteigt. Lavine et al. (1998) bilden einen Index der Einstellungsstärke aus den sechs Attributen „Wichtigkeit“, „Sicherheit“, „Intensität“, „Häufigkeit des Nachdenkens“, „Extremität“ und „Ambivalenz“. Während die einzelnen Attribute keine moderierende Wirkung auf das Auftreten von Fragereihenfolgeeffekten nehmen, tritt ein Index aller sechs Attribute hingegen als signifikante Kovariate in der Analyse auf. Dieses Ergebnis ist allerdings nicht für alle drei durchgeführten Experimente zu beobachten. Demgegenüber stehen auch Studien, die keine moderierende Rolle der Einstellungsstärke auf das Auftreten von Kontexteffekten beobachten (Bassili und Krosnick 2000; Bishop 1990). Als Erklärung dieses Puzzles aus verschiedenen Ergebnissen wird angeführt, dass jede Art von Kontexteffekt einen anderen Ursprung im kognitiven Prozess der Antwortabgabe besitzt und daher Einstellungsstärke bzw. ihre einzelnen Attribute nicht alle Effekte gleichermaßen moderieren (Bassili und Krosnick 2000). Studien, die das Auftreten von Effekten nonverbaler, graphischer Ratingskalenelemente in

Abhängigkeit von der Höhe der Einstellungsstärke untersuchen, wurden (meines Wissens nach) bisher nicht veröffentlicht.

2.2.2 Eigenschaften der Ratingskala

Hierarchie verbaler und nonverbaler Fragebogenelemente

Befragte ziehen alle im Erhebungsinstrument gegebenen Informationen heran, um ihre Aufgabe, die Beantwortung der Fragen, zu erfüllen. Die Frage ist, ob alle Arten von Informationen - verbale, numerische, symbolische und graphische - dabei als gleichwertig betrachtet werden, oder ob sie einer Hierarchie unterliegen. Empirisch zeigt sich, dass Befragte eher auf die subtilere, graphische Information von Antwortskalen reagieren, wenn verbale und numerische Information lediglich in geringem Umfang vorliegen, wie beispielsweise in Ratingskalen die lediglich an den Endpolen mit Etiketten versehen sind (Toepoel und Dillman 2011; Tourangeau et al. 2007).

Darstellung der Items

Da in der Regel eine Hierarchie zwischen den Arten der Information in Ratingskalen vorliegt, bei der die verbalen und numerischen Elemente den graphischen überlegen sind, scheint die Darstellungsform der Items eine mögliche weitere Einflussgröße für das Auftreten von Effekten des Ratingskalenlayouts zu sein. Vor allem die Einzel-Item-Darstellung unterscheidet sich stark von einer Darstellung der Items in einer Batterie. Während in der Einzel-Item-Darstellung auf das Item direkt die Ratingskala folgt, sind in der Item-Batterie in der Regel mehrere Items tabellarisch untereinander aufgelistet und die Antwortmöglichkeiten finden sich rechts nebenstehend. Bei der Einzel-Item-Darstellung stehen (sofern vorhanden) die verbalen Etiketten direkt über den Antwortkästchen, was in der Item-Batterie nicht der Fall ist. Hier stehen die verbalen Etiketten meist in der oberen Reihe und in den folgenden Reihen sind dann lediglich die Antwortkästchen abgebildet. In der Item-Batterie ist also viel weniger verbale Information enthalten als in einer Einzel-Item-Darstellung. Gemäß der zuvor beschriebenen Hierarchie der Informationsarten, sind Effekte der nonverbalen Skalenelemente in Item-Batterien wahrscheinlicher als in der Einzel-Item-Darstellung. Diese Annahme wurde bisher empirisch nicht untersucht. Zwar wird in dieser Dissertation erstmals eine

Itematterie eingesetzt, um Layouteffekte zu untersuchen, aber ein systematischer Vergleich mit einer Einzel-Item-Darstellung wird nicht vorgenommen.

3. Don't know-Kategorie in Ratingskalen

Die Vorgabe von Don't know-Kategorien in Ratingskalen ist ein viel diskutiertes Thema der methodischen Umfrageforschung. Im folgenden Unterabschnitt (3.1) wird der Forschungsstand anhand dieser Diskussion vorgestellt. Dabei wird der Begriff „Don't know-Kategorie“ (DK-Kategorie) als Synonym für die verschiedenen nicht-inhaltlichen Antwortkategorien, wie beispielsweise „keine Meinung“, „weiß-nicht“ oder „kann ich nicht sagen“, benutzt. Analog dazu umfasst der Begriff „Don't know-Antwort“ (DK-Antwort) Antworten zu allen Arten nicht-inhaltlicher Antwortkategorien und mit dem Begriff „Don't know-Vorgabe“ (DK-Vorgabe) ist schließlich die Vorgabe aller nicht-inhaltlichen Kategorien gemeint. Der zweite Unterabschnitt (3.2) stellt die Survey Satisficing-Theorie nach Krosnick (Krosnick 1991) vor. Sie dient als Erklärungsmodell dafür, warum auch Befragte, die in Besitz einer relevanten Einstellung sind, von der DK-Kategorie Gebrauch machen. Unterabschnitt (3.3) gibt einen Überblick über die verschiedenen Möglichkeiten der DK-Vorgabe in den unterschiedlichen Erhebungsmodi. Zudem werden hier Unterschiede der Modi hinsichtlich der Häufigkeiten von DK-Antworten behandelt und Erklärungsansätze dazu geboten. Unterabschnitt (3.4) stellt den Forschungsstand zu Kovariate von DK-Antworten vor und in Unterabschnitt (3.5) werden als Fazit aus den vorangehenden Unterabschnitten dieses dritten Abschnitts Empfehlungen hinsichtlich der Verwendung von DK-Kategorien formuliert.¹

3.1 Diskussion der Don't know-Kategorie in Ratingskalen

Die offensichtliche Intention der Vorgabe von DK-Kategorien in Ratingskalen ist, Befragten, die keine Einstellung gegenüber dem Frageinhalt besitzen, die

¹ Die Ausführungen dieses Abschnitts basieren zum Teil auf der Publikation von Menold und Bogner (2014) und Bogner und Landrock (2014).

Möglichkeit zu bieten, dies auch wahrheitsgemäß zu berichten. Wird hingegen keine DK-Kategorie angeboten, sehen sich diese Befragten dazu gedrängt, zufällig aus den inhaltlichen Antwortalternativen zu wählen, was als non-attitude reporting bezeichnet wird (Converse 2006). Diese zufälligen Antworten erhöhen die Fehlervarianzen und den Messfehler, was zu einer verminderten Datenqualität führt (Andrews 1984). Deshalb empfehlen Vertreter der klassischen Position die Vorgabe einer DK-Kategorie in Ratingskalen (vgl. u.a. Converse und Presser 1986; Katz 1942; Payne 1950; Schuman und Presser 1981; Vaillancourt 1973).

Dieser Argumentation, die für eine Verwendung von DK-Kategorien spricht, mit dem Ziel, Non-attitude Reporting zu vermeiden und die Datenqualität zu steigern, liegen zwei Annahmen zugrunde (Krosnick 2002). Die erste ist, dass es lediglich zwei Typen von Befragten gibt: nämlich solche, die eine Einstellung zu dem Thema besitzen und sich darüber auch im Klaren sind und solche, die eben keine Einstellung besitzen und denen dies auch bewusst ist. Diese Annahme impliziert, dass Befragte über einen hohen Grad an Wissen über sich selbst verfügen. Die zweite Annahme lautet, dass Befragte rational handeln. Treffen diese beiden Annahmen zu, dann wird vom ersten Befragtentyp erwartet, dass er seine Einstellung berichtet, unabhängig davon, ob eine DK-Kategorie angeboten wird oder nicht. Die Erwartung hinsichtlich des Verhaltens des zweiten Befragtentyps ist, dass er die DK-Kategorie wählt, sofern sie denn vorgegeben ist. Sollte dies nicht der Fall sein, wird dieser Typ, wie bereits zuvor erwähnt, wahrscheinlich eine Einstellung durch die Wahl einer der inhaltlichen Antwortalternativen vortäuschen. Damit vermeidet der Befragte, meinungslos zu erscheinen. Denn aufgrund der Tatsache, dass keine DK-Kategorie vorgegeben wird, scheint in diesem Kontext Meinungslosigkeit keine legitime Option darzustellen.

Vertreter der neueren Position weisen jedoch darauf hin, dass die Verwendung von DK-Kategorien insoweit problematisch ist, als dass nicht nur befragte Personen ohne, sondern auch solche mit einer relevanten Einstellung zu dem abgefragten Thema diese Kategorie zur Antwortabgabe heranziehen (vgl. u.a. Gilljam und Granberg 1993; Krosnick und Fabrigar 1997). So weichen Befragte auf die DK-Kategorie aus, wenn sie unsicher hinsichtlich der Bedeutung der Frage (Feick 1989) oder unsicher hinsichtlich der Bedeutung der Antwortformate (Krosnick und Fabrigar 1997) sind. Auch antworten befragte Personen mittels der DK-Kategorie,

wenn sie miteinander in Konflikt stehende Einstellungen besitzen (Coombs und Coombs 1976). Befragte können mittels einer DK-Antwort auch die Äußerung von nicht sozial erwünschten Einstellungen vermeiden (Krosnick und Fabrigar 1997). Außerdem kann das Vorliegen einer DK-Kategorie von den Befragten dahingehend interpretiert werden, dass elaboriertes Wissen zur Beantwortung der Frage erforderlich ist, was zu Unsicherheit und somit zu einem Ausweichen auf die DK-Kategorie führen kann (Hippler und Schwarz 1989). Schließlich zeigen weitere Studien, dass Befragte eine nicht-inhaltliche Antwort geben, um intensives Nachdenken zu vermeiden bzw. um sich nicht auf die Befragung einzulassen (Oppenheim 1992), oder wenn ihre Motivation oder ihre kognitiven Fähigkeiten zu gering sind, um die Fragen gründlich zu beantworten (Krosnick, 1991).

Hinsichtlich der Datenqualität kommen verschiedene Studien zu dem Ergebnis, dass der Ausschluss von DK-Kategorien diese nicht negativ beeinflusst (vgl. u.a. Alwin und Krosnick 1991; Krosnick et al. 2002; McClendon und Alwin 1993; Poe et al. 1988). In einer Wahlstudie wurden schließlich exaktere Wahlprognosen erzielt, wenn Befragte, die eine DK-Antwort gewählt hatten, anschließend unter Nachdruck um eine inhaltliche Antwort gebeten wurden (Visser et al. 2000). Weitere Studien finden, dass die Vorgabe von DK-Kategorien keinen signifikant positiven Effekt auf die Validität und Reliabilität der Frage ausübt (Scherpenzeel und Saris 1997).

Wird allerdings keine DK-Kategorie angeboten, nutzen Befragte häufig die Mittelkategorie als Ersatz, was jedoch keinen Einfluss auf die Validität und Reliabilität der von Kulas et al. (2008) untersuchten Persönlichkeitsskalen nimmt. Sturgis et al. (2014) beobachten in ihrer Studie, dass Befragte ohne Einstellung die Mittelkategorie wählen, mit der Intention, ihre Meinungslosigkeit verbergen zu wollen. Anhand offener Nachfragen zeigen sie in ihrer Studie, dass vor allem solche Befragte ohne Einstellung, die der Meinung sind, dass sie zu wichtigen Themen eine Einstellung besitzen und berichten sollten, Mittelkategorie-Antworten geben. Dadurch wird ein systematischer Fehler in die Daten eingeführt. Werden diese Mittelkategorie-Antworten als DK-Antworten behandelt und von der Analyse ausgeschlossen, ändert sich die Verteilung der untersuchten Items signifikant (Sturgis et al. 2014).

3.2 Survey Satisficing-Theorie

Die Survey Satisficing-Theorie nennt geringe Motivation und geringe kognitive Fähigkeiten als Erklärungsfaktoren für DK-Antworten von Befragten, die eigentlich eine relevante Einstellung besitzen. Nach dieser Theorie existieren zwei alternative Tracks des kognitiven Fragebeantwortungsprozesses (Krosnick 1991,1999; Krosnick et al. 1996). Generell geht die Fragebeantwortung in Umfragen mit einer kognitiven Belastung für den Befragten einher: Befragte müssen (1.) die Frage verstehen, (2.) in ihrem Gedächtnis nach relevanten Informationen suchen und (3.) daraus eine Beurteilung ihrer persönlichen Einstellung generieren, die sie schließlich (4.) anhand einer der angebotenen Antwortvorgaben berichten (Tourangeau et al. 2000). Im ersten, als Optimizing bezeichneten Track durchlaufen Befragte diese vier kognitiven Prozesse in gründlicher und unvoreingenommener Weise und berichten ihre wahre Antwort. Im zweiten, als Satisficing bezeichneten Track richten die Befragten ihr Verhalten auf die Reduzierung der kognitiven Belastung aus, in dem sie kleinere oder größere Vereinfachungen des kognitiven Prozesses vornehmen. Während bei einem schwachen Satisficing noch alle vier kognitiven Schritte durchlaufen werden - wenn auch weniger gewissenhaft als beim Optimizing - wird bei stärkerem Satisficing der Informationsverarbeitungsprozess stärker verkürzt bzw. bei sehr starkem Satisficing überhaupt nicht vollzogen. Beispielsweise wird die Frage gar nicht gelesen. Befragte, die ein Satisficing-Verhalten zeigen, beurteilen die Fragestellung lediglich oberflächlich und berichten die Antwort, die ihnen zuerst in den Sinn kommt oder die ihnen als plausibel erscheint (Krosnick et al. 1996). Im Wesentlichen bestimmen drei Faktoren die Wahrscheinlichkeit, dass ein Befragter bei der Fragebeantwortung eine Satisficing-Strategie wählt: (1.) die Schwierigkeit der Aufgabe, (2.) seine kognitiven Fähigkeiten und (3.) seine Motivation, die Aufgabe zu erfüllen (Krosnick et al. 1996; Krosnick und Presser 2010). Nach der Survey Satisficing-Theorie stellen explizit angebotene DK-Kategorien für Befragte, die ein Satisficing-Verhalten zeigen, eine einfache und legitime Möglichkeit zur Reduzierung der kognitiven Belastung des Fragebeantwortungsprozesses dar.

3.3 Don't know-Vorgabe in verschiedenen Erhebungsmodi

Der jeweilige Erhebungsmodus bestimmt über die Möglichkeiten der Vorgabe von DK-Kategorien. In selbstadministrierten Umfragen, die anhand von Papierfragebögen erhoben werden, muss bereits bei der Entwicklung des Fragebogens entschieden werden, ob eine DK-Kategorie explizit vorgegeben wird oder nicht. Wird die DK-Kategorie nicht explizit vorgegeben, werden DK-Antworten damit mit einer, beispielsweise aus Gründen der Verweigerung oder der Unachtsamkeit, ausgelassenen Antwort gleichgesetzt.

In Face-to-Face- oder telefonischen Interviews besteht hingegen die Wahl zwischen der expliziten und der impliziten Vorgabe von DK-Kategorien. Anders als in selbstadministrierten, papierbasierten Befragungen kann in Interviews bei keiner Vorgabe einer DK-Kategorie eine vom Befragten selbständig geäußerte, fehlende Einstellung als DK-Antwort durch den Interviewer akzeptiert werden. Dies wird als implizite DK-Vorgabe bezeichnet. Oftmals gilt in diesem Fall für Interviewer die Anweisung, bei einer selbständig geäußerten DK-Antwort einmalig nachzufragen, ob nicht doch eine Einstellung vorliegt, dann aber die selbständig geäußerte DK-Antwort als solche zu notieren. Selbständig geäußerte DK-Antworten unterscheiden sich von Antworten zu vorgegebenen DK-Kategorien dahingehend, als dass es sich bei ihnen um vom Befragten selbständig konstruierte Antworten innerhalb des Befragungsprozesses handelt. Das Vorgehen der impliziten DK-Vorgabe stellt eine Möglichkeit dar, die Legitimität einer DK-Antwort dem Befragten nicht offensichtlich zu unterbreiten (Dillman und Christian 2005).

Selbstadministrierte Befragungen, die mittels interaktiver Fragebögen am Computer erhoben werden (z.B. Web Surveys), bieten eine Reihe verschiedener technischer Umsetzungsmöglichkeiten der DK-Vorgabe: Erstens kann eine DK-Kategorie explizit angeboten werden. Zweitens kann auf die Vorgabe einer DK-Kategorie verzichtet, aber bei Nichtbeantwortung der Frage mittels Follow-up Probes darum gebeten werden, doch eine inhaltliche Antwort zu geben oder eine DK-Antwort zu bestätigen. Dieses Vorgehen gleicht der impliziten DK-Vorgabe in Interviews. Bei einem Vergleich dieser beider Umsetzungsmöglichkeiten finden Derouvray und Couper (2002) die geringste Rate an Item-Nonresponse für das zweite Design, in dem die DK-Kategorie erst im Follow-up Probe vorgegeben wird. Darüber hinaus gibt es weitere Kombinationsmöglichkeiten der expliziten oder impliziten DK-

Vorgabe mit Follow-up Probes, die alle darauf zielen, die Rate an Item-Nonresponse so gering wie möglich zu halten.

Verschiedene Untersuchungen kommen zu dem Ergebnis, dass sich die Anteile an DK-Antworten signifikant zwischen den verschiedenen Erhebungsmodi unterscheiden. Dabei zeigt sich eine signifikant niedrigere Rate an DK-Antworten für Face-to-Face-Interviews im Vergleich zu selbstadministrierten, papierbasierten Umfragen (De Leeuw 1992), im Vergleich zu Web Surveys (Heerwegh 2009; Heerwegh und Loosveldt 2008) und im Vergleich zu Telefoninterviews (Holbrook et al. 2003). Des Weiteren wählen Befragte signifikant häufiger die DK-Kategorie in Web Surveys als in Telefoninterviews (Fricker et al. 2005).

Holbrook et al. (2003) argumentieren, dass lediglich in Face-to-Face-Interviews eine nonverbale Kommunikation zwischen Interviewer und Befragtem stattfindet. Von dieser nonverbalen Kommunikation wird angenommen, dass sie zusätzlich die Motivation des Befragten erhöhen kann und somit der Tendenz zu Satisficing und damit der Wahrscheinlichkeit von DK-Antworten entgegenwirkt. Ein zweites Argument dafür, warum die Rate an DK-Antworten in Face-to-Face-Interviews geringer ist als in anderen Erhebungsmodi, lautet, dass nur in diesem Erhebungsmodus gesichert ist, dass sich der Befragte voll und ganz auf die Befragung konzentriert. In Telefoninterviews und selbstadministrierten Befragungen können Befragte neben der Fragebeantwortung gleichzeitig auch noch in weitere, andersartige Aufgaben involviert sein, was sie in ihrer Konzentration stören kann (Holbrook et al. 2003). Schließlich unterscheidet sich als dritter Punkt auch die Aufgabenschwierigkeit zwischen den einzelnen Modi, wobei sie für Face-to-Face-Interviews niedriger eingestuft wird als für selbstadministrierte Umfragen. In Face-to-Face-Interviews bekommen die Befragten die Frage vorgelesen und berichten ihre Antwort mündlich. Anderweitige Aufgaben, wie beispielsweise das selbständige Folgen von Instruktionen oder die eigenständige Bedienung des computerisierten Befragungsinstruments müssen sie, anders als Befragte in selbstadministrierten Befragungen, nicht erfüllen (Bowling 2005). Befragte in Face-to-Face-Interviews wählen daher die DK-Kategorie seltener als in anderen Erhebungsmodi, weil sie eine niedrigere Belastung verspüren und diese demnach auch weniger reduzieren müssen (Heerwegh und Loosveldt 2008).

3.4 Kovariate von Don't know-Antworten

Empirische Untersuchungen belegen, dass verschiedene sozio-demographische Gruppen und bestimmte Bedingungen der Befragungssituation und des Befragungsinstrumentes die Wahl von DK-Kategorien begünstigen. So wählen Befragte mit einer niedrigen formalen Bildung wahrscheinlicher die DK-Kategorie als höher gebildete Befragte (Faulkenberry 1978; Krosnick 1999; Krosnick et al. 2002). Auch ist ein Geschlechterunterschied zu erkennen: Frauen wählen häufiger die DK-Kategorie zur Antwortabgabe als Männer, was nach Rapoport (Rapoport 1982,1985) damit erklärt werden kann, dass Frauen aufgrund ihrer Sozialisation eine stärkere Unsicherheit dabei verspüren, ihre Meinung öffentlich zum Ausdruck zu bringen als dies bei Männern der Fall ist. Auch Befragte, die angeben, sich bei der Beantwortung der Fragen nicht besonders angestrengt zu haben, - ein Indikator für Satisficing - wählen die DK-Kategorie wahrscheinlicher als solche, die mehr Anstrengung im Befragungsprozess eingebracht haben. Des Weiteren steigt die Wahrscheinlichkeit von DK-Antworten im späteren Verlauf der Befragung, also dann, wenn die Motivation für ein Optimizing-Verhalten wahrscheinlich zurückgeht (Ferber 1966; Krosnick et al. 2002). Es werden mehr DK-Antworten beobachtet, wenn Antworten offen gegeben werden als wenn eine geheime Antwortabgabe möglich ist (Krosnick et al. 2002). Schließlich wird bei Fragen, die sensible Themen betreffen, häufiger die DK-Kategorie gewählt, um Äußerungen von sozial unerwünschten Antworten zu vermeiden bzw. aus Gründen des Selbstschutzes (Johanson et al. 1993).

3.5 Empfehlungen für die Praxis

Aus dem in Unterabschnitt (3.1) beschriebenen Forschungsstand lässt sich eine generelle Empfehlung für oder wider die Verwendung von DK-Kategorien nicht aussprechen. Viele Studien kommen zwar zu dem Schluss, dass die Vorgabe von DK-Kategorien nicht sinnvoll ist, weil sie nicht nur Befragte ohne, sondern auch mit einer relevanten Einstellung bzw. dem nötigen Wissen zur Beantwortung der Frage nutzen (vgl. u.a. Alwin und Krosnick 1991; Krosnick et al. 2002). Allerdings kontrollieren die Studien nicht weitere Einflussgrößen auf DK-Antworten, wie Frageinhalt oder Attribute der Antwortskala, wie z.B. die Anzahl der Kategorien

oder die Verwendung verbaler und numerischer Labels. Zudem finden andere Studien, dass sich die Datenqualität durch die Vorgabe von DK-Kategorien erhöhen lässt (vgl. u.a. Andrews 1984).

Daher muss jede Forscherin / jeder Forscher für ihre / seine Umfrage entscheiden, ob DK-Kategorien angeboten werden oder nicht. Diese Entscheidung sollte in Abhängigkeit verschiedener Faktoren getroffen werden:

Erstens sollen der Frageinhalt und der Wissenstand der Zielgruppe berücksichtigt werden. Bei Faktfragen, die objektive Informationen über die Befragten erheben, wie beispielsweise soziodemographische Merkmale oder die Ausstattung des Haushalts, ist eine DK-Kategorie nicht sinnvoll. Hier wird davon ausgegangen, dass die befragten Personen die Antwort kennen. Bei Fragen, die den Wissensstand der Befragungspopulation zu einem bestimmten Thema erheben, stellt die DK-Kategorie hingegen eine legitime Antwortoption dar und sollte angeboten werden. Schließlich muss bei Fragen zu Einstellungen überlegt werden, ob die untersuchte Zielgruppe über eine relevante Einstellung verfügt, was wohl bei vielen Themen der Fall sein dürfte.

Zweitens spielt bei der Entscheidung der Vorgabe von DK-Kategorien auch der Erhebungsmodus eine Rolle. Außer in selbstadministrierten, papierbasierten Umfragen bieten sich neben einer expliziten DK-Vorgabe weitere Darbietungsmöglichkeiten. Sowohl bei Antworten zu expliziten DK-Kategorien als auch bei selbständig geäußerten DK-Antworten können Follow-up Probes, also Nachfragen, folgen. Dies kann je nach Modus durch den Interviewer oder automatisiert durch das computerisierte Befragungsinstrument geschehen. Durch solche Follow-up Probes sollen Befragte doch noch zum Berichten einer inhaltlichen Antwort gedrängt werden bzw. die Information über eine mögliche Tendenz erfragt werden (Bradburn und Sudman 1988). Jedoch empfehlen Sanchez und Morchio (1992), in Interviews nur bei Einstellungsfragen Follow-up Probes zu verwenden. Ihre Untersuchung von Wissensfragen zeigt, dass die Befragten bei Nachfragen dazu tendieren, eine Antwort zu raten, da sie über kein Wissen verfügen, auf das sie eine Antwort basieren können. Als Folge erhöht sich der systematische Messfehler in den Daten.

Drittens hängt die Entscheidung für oder wider die DK-Kategorie auch von der Stichprobengröße ab. Durch vorgegebene DK-Kategorien wird meist die effektive Stichprobengröße reduziert. Da DK-Antworten in der Regel von der Datenanalyse ausgeschlossen werden, kann dies bei kleinen effektiven Stichproben zu geringen Fallzahlen führen, die für fortgeschrittene Analyseverfahren nicht mehr ausreichend sind.²

4. Effekte des Layouts der Don't know-Kategorie in Ratingskalen

Ist eine explizite Vorgabe einer DK-Kategorie in einer Ratingskala vorgesehen, ist für visuell präsentierte Befragungsinstrumente, beispielsweise Web Surveys oder Papierfragebögen, über das Layout der DK-Kategorie zu entscheiden. Aufgrund des großen gestalterischen Spielraums existiert eine Vielzahl an möglichen Layouts der DK-Kategorie. Allerdings haben sich in sozialwissenschaftlichen Umfragen vor allem zwei Layouts durchgesetzt. Im ersten Layout wird die DK-Kategorie mittels eines Trennstriches bzw. mittels eines Leerraums von den inhaltlichen Antwortkategorien der Ratingskala abgetrennt (SEP-DK-Layout). Im zweiten Layout wird die DK-Kategorie einfach an die inhaltlichen Kategorien angehängt (ADD-DK-Layout) (vgl. Abbildung 1).

Bisher thematisieren die empirischen Studien von Tourangeau et al. (2004) und Christian et al. (2009) den Einfluss der graphischen Darstellung der DK-Kategorie in Ratingskalen in Form des SEP-DK- und des ADD-DK-Layouts auf das Antwortverhalten von Befragten. Basierend auf der Heuristik „middle means typical“ wurde die Hypothese aufgestellt, dass bei einer Verschiebung des visuellen

² An dieser Stelle soll kurz erwähnt werden, dass Filterfragen eine Alternative zur Verwendung von DK-Kategorien bieten. Filterfragen erheben, ob die befragte Person eine relevante Einstellung besitzt. Ist dies der Fall, wird diese Einstellung detailliert abgefragt, anderenfalls wird die nächste Frage gestellt. Ziel dieser Filterung ist, befragte Personen ohne Einstellung nicht zu einer falschen inhaltlichen Antwort zu drängen und somit die Datenqualität zu verbessern. Der Nachteil von Filterfragen ist allerdings, dass die Rate an DK-Antworten deutlich höher liegt als bei Ratingskalen mit DK-Kategorie (Schuman & Presser, 1981). Zudem erhöht eine Formulierung des Filters, die eine intensive Auseinandersetzung mit dem Thema impliziert, die Wahrscheinlichkeit einer Verneinung einer existierenden Einstellung (vgl. u.a. Bishop et al. 1983; Fowler und Cannell 1996; Hippler und Schwarz 1989; Krosnick und Abelson 1991). Dieser Effekt der Filterfrageformulierung wird umso stärker, je abstrakter oder weniger bekannt das inhaltliche Thema der Frage ist (Bishop et al. 1983).

Mittelpunktes weg von der inhaltlichen Mitte der Skala, wie es für das ADD-DK-Layout zutrifft, eine Verschiebung der Antworten in Richtung des visuellen Mittelpunktes verursacht (Tourangeau et al. 2004).

SEP-DK-Layout		Abtrennung mittels eines Trennstrichs			
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	DK
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SEP-DK-Layout		Abtrennung mittels eines Leerraums			
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	DK
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ADD-DK-Layout		Keine graphische Abtrennung			
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	DK
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 1 Drei Varianten der graphischen Darstellung der DK-Kategorie einer Ratingskala

In den beiden folgenden Unterabschnitten (4.1) und (4.2) werden diese Studien im Detail vorgestellt. Da sie zu konträren Ergebnissen gelangen, werden schließlich in Unterabschnitt (4.3) mögliche Ursachen dafür diskutiert.

4.1 Die Studie von Tourangeau, Couper und Conrad

Tourangeau et al. (2004) testen die Hypothese mittels eines Split-Ballot-Experiments in zwei Web Surveys in den Jahren 2001 und 2002. Im ersten Web Survey-Experiment bekommen 2.987 Befragte eines der beiden Layouts zufällig vorgegeben. Im zweiten Web Survey-Experiment wird ein drittes Layout, in dem die Trennung anstatt mittels eines Trennstrichs mittels eines Leerraums vollzogen wird, hinzugenommen. Die drei Layouts werden über alle 1.590 Befragten der zweiten Web Survey-Studie randomisiert. Als Antwortskala wird eine vertikale, fünfstufige, vollverbalisierte bipolare Ratingskala (vgl. Abbildung 2, Seite 33) mit zwei nicht-inhaltlichen Kategorien („Don’t know“; „No opinion“) vorgegeben. In

beiden Web Survey-Experimenten werden jeweils die folgenden zwei Items eingesetzt, zu deren Beantwortung die variierten Skalenlayouts vorgegeben werden:

Item 1 *“Think of how much the federal government is doing to make sure women have the same job opportunities as men. Would you say the federal government is doing too much, about the right amount, or too little about this?”* (Tourangeau et al. 2004, S.375)

Item 2 *“Think of how much the federal government is doing to provide day care centers for the children of working parents. Would you say the federal government is doing too much, about the right amount, or too little about this?”* (Tourangeau et al. 2004, S.375)

Die Ergebnisse beider Web Survey-Experimente zeigen Unterschiede in den Antwortverteilungen des SEP-DK-Layouts, in dem die inhaltliche und die visuelle Mitte übereinstimmen, und des ADD-DK-Layouts, in dem die Mittelpunkte voneinander abweichen. Da sich die Ergebnisse der SEP-DK-Layouts mit Trennstrich bzw. Leerraum im zweiten Web Survey-Experiment nicht wesentlich voneinander unterscheiden, werden sie von den Autoren zusammengefasst betrachtet. In allen vier untersuchten Vergleichen zeigen sich höhere Mittelwerte, wenn die nicht-inhaltlichen Kategorien an die inhaltlichen angehängt (ADD-DK-Layout) sind, anstatt von ihnen getrennt zu sein (SEP-DK-Layout). Drei dieser vier Mittelwertunterschiede sind statistisch signifikant. Dieses Ergebnis unterstützt die Hypothese und zeigt, dass Befragte durch den verschobenen visuellen Mittelpunkt im ADD-DK-Layout in der Skaleninterpretation fehlgeleitet werden können und anscheinend eher den visuellen als den inhaltlichen Skalenmittelpunkt als Referenzpunkt für ihre Antwortabgabe nutzen (Tourangeau et al. 2004). In den Daten zeigte sich jedoch ein zusätzlicher Trend: In einem der vier experimentellen Vergleiche wählen Befragte signifikant häufiger eine der nicht-inhaltlichen Kategorien, wenn diese mittels eines Trennstriches von der inhaltlichen Ratingskala abgetrennt sind. Es hat den Anschein, dass die graphische Abtrennung die nicht-inhaltliche Kategorie visuell hervorhebt und die Aufmerksamkeit der Befragten auf sie lenkt.

4.2 Die Studie von Christian, Parsons und Dillman

Christian et al. (2009) führen in einem Studenten-Web-Survey im Jahr 2003 eine ähnliche Studie durch, mit dem Ziel die Ergebnisse von Tourangeau et al. (2004) zu replizieren. Es werden wieder zwei Layouts in einem Split-Ballot-Experiment getestet: Im ersten Layout wird die nicht-inhaltliche Antwortkategorie anhand eines breiteren Leerraums von den inhaltlichen abgetrennt (SEP-DK-Layout) und im zweiten einfach angehängt (ADD-DK-Layout). Die Layouts werden über 1.565 Befragten randomisiert. Die Antwortskala ist eine vertikale, fünfstufige, vollverbalisierte, bipolare Skala mit einer nicht-inhaltlichen Kategorie (Don't know). Die inhaltlichen Kategorien sind zusätzlich noch mit numerischen Etiketten von eins bis fünf versehen (vgl. Abbildung 2, Seite 33). Das Experiment wird anhand zweier Items erhoben, wobei aus der veröffentlichten Literatur nur der Fragetext eines Items bekannt ist:

Item 1 *“On a scale of 1 to 5, where 1 means very satisfied and 5 means very dissatisfied, how satisfied are you with the variety of cultural activities offered at the Washington State University?”* (Christian et al. 2009, S.405)

Konträr zu den Ergebnissen der früheren Studie werden im SEP-DK-Layout, in dem inhaltliche und visuelle Mitte übereinstimmen, höhere Mittelwerte beobachtet als im ADD-DK-Layout, für welches die Verschiebung der Antwortverteilung Richtung Skalenende eigentlich erwartet worden war. Auch wählen Befragte die DK-Kategorie nicht häufiger, wenn sie abgetrennt von den inhaltlichen Kategorien dargestellt und damit visuell hervorgehoben wird. Lediglich die stärkere Tendenz zur Mittelkategorie, wenn diese mit der visuellen Mitte übereinstimmt, kann in dieser Studie repliziert werden. Eine Paradataanalyse bringt schließlich noch das Ergebnis, dass Befragte im SEP-DK-Layout im Mittel länger zur Beantwortung der Items benötigen als Befragte im ADD-DK-Layout. Dies deutet darauf hin, dass Befragte länger benötigen, um alle Information im SEP-DK-Layout kognitiv zu verarbeiten (Christian et al. 2009).

4.3 Mögliche Ursachen der konträren Ergebnisse beider Studien

In diesen beiden vorgestellten Studien wird die gleiche Hypothese untersucht: Wenn sich Befragte bei ihrer Antwortabgabe stärker an der visuellen als der inhaltlichen Mitte einer Skala orientieren, dann hat eine unterschiedliche graphische Darstellung der DK-Kategorie und die damit einhergehende Verschiebung des visuellen Mittelpunktes einen Einfluss auf die Antwortverteilung. Ein direkter Vergleich der Ergebnisse der beiden Studien von Tourangeau et al. (2004) und Christian et al. (2009) ist allerdings aufgrund folgender Unterschiede im Studiendesign nicht einfach möglich:

1. Anzahl der nicht-inhaltlichen Kategorien

In der Studie von Tourangeau et al. (2004) werden zwei DK-Kategorien an die fünfstufige Ratingskala angehängt, weshalb sich der visuelle Mittelpunkt im ADD-DK-Layout um eine ganze Antwortkategorie hin zum Skalenende auf die Kategorie „too little“ verschiebt. In der Studie von Christian et al. (2009) wird lediglich eine DK-Kategorie verwendet. Dadurch fällt die visuelle Verschiebung geringer aus, der visuelle Mittelpunkt liegt zwischen der Mittelkategorie und der Kategorie „somewhat dissatisfied“ (vgl. Abbildung 2).

SEP-DK-Layout			
Tourangeau et al. (2004)		Christian et al. (2009)	
<input type="radio"/> Far too much		<input type="radio"/> 1 Very satisfied	
<input type="radio"/> Too much		<input type="radio"/> 2 Somewhat satisfied	
<input type="radio"/> About the right amount		<input type="radio"/> 3 Neither satisfied nor dissatisfied	
<input type="radio"/> Too little		<input type="radio"/> 4 Somewhat dissatisfied	
<input type="radio"/> Far too little		<input type="radio"/> 5 Very dissatisfied	
<hr/>			
<input type="radio"/> Don't know		<input type="radio"/> Don't know	
<input type="radio"/> No opinion			
ADD-DK-Layout			
Tourangeau et al. (2004)		Christian et al. (2009)	
<input type="radio"/> Far too much		<input type="radio"/> 1 Very satisfied	
<input type="radio"/> Too much		<input type="radio"/> 2 Somewhat satisfied	
<input type="radio"/> About the right amount		<input type="radio"/> 3 Neither satisfied nor dissatisfied	
<input type="radio"/> Too little		<input type="radio"/> 4 Somewhat dissatisfied	
<input type="radio"/> Far too little		<input type="radio"/> 5 Very dissatisfied	
<input type="radio"/> Don't know		<input type="radio"/> Don't know	
<input type="radio"/> No opinion			

Abbildung 2 SEP-DK- und ADD-DK-Layout der Ratingskalen bei Tourangeau et al. (2004) und Christian et al. (2009)

2. *Verbale Etiketten*

Während Tourangeau et al. (2004) in ihrer Studie die verbalen Etiketten „far too much“ bis „far too little“ verwenden, sind es in der Studie von Christian et al. (2009) verbale Etiketten von „very satisfied“ bis „very dissatisfied“ (vgl. Abbildung 2).

3. *Nummerische Etiketten*

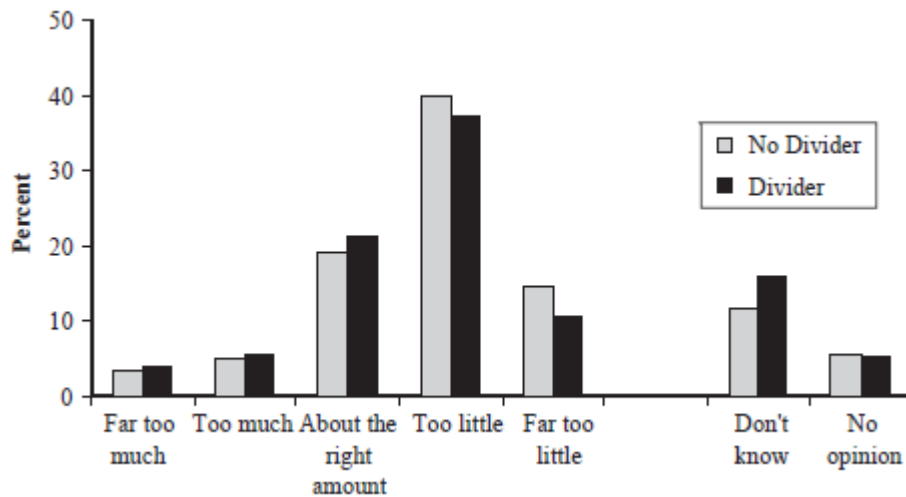
In der Studie von Christian et al. (2009) sind auch die inhaltlichen Kategorien zusätzlich zu den verbalen mit numerischen Etiketten von eins bis fünf versehen, die DK-Kategorie hingegen nicht. Daher unterscheidet sich die DK-Kategorie im ADD-DK-Layout durch die fehlende Nummer von den inhaltlichen, auch wenn sie nicht graphisch abgetrennt wurde (vgl. Abbildung 2).

4. *Thema des Items*

In der Studie von Tourangeau et al. (2004) werden Befragte um eine Einschätzung gebeten, ob die Regierung zu viel oder zu wenig unternimmt, um Chancengleichheit im Beruf bzw. eine ausreichende Kinderbetreuung für Erwerbstätige zu gewährleisten. Ein abstraktes, politisches Thema, zu dem sicherlich nicht alle Befragten eine vorgefertigte Meinung abrufbereit im Gedächtnis vorliegen haben. Befragte müssen den kognitiven Prozess der Fragebeantwortung gründlich durchlaufen, um ihre tatsächliche Einstellung berichten zu können. Christian et al. (2009) befragen in ihrer Studie Studenten über ihre Zufriedenheit bzw. Unzufriedenheit mit dem kulturellen Angebot auf dem Universitätscampus. Ein Thema aus dem Lebensalltag der Studenten, dessen Beantwortung in den meisten Fällen wohl einen eher geringen kognitiven Aufwand erfordert.

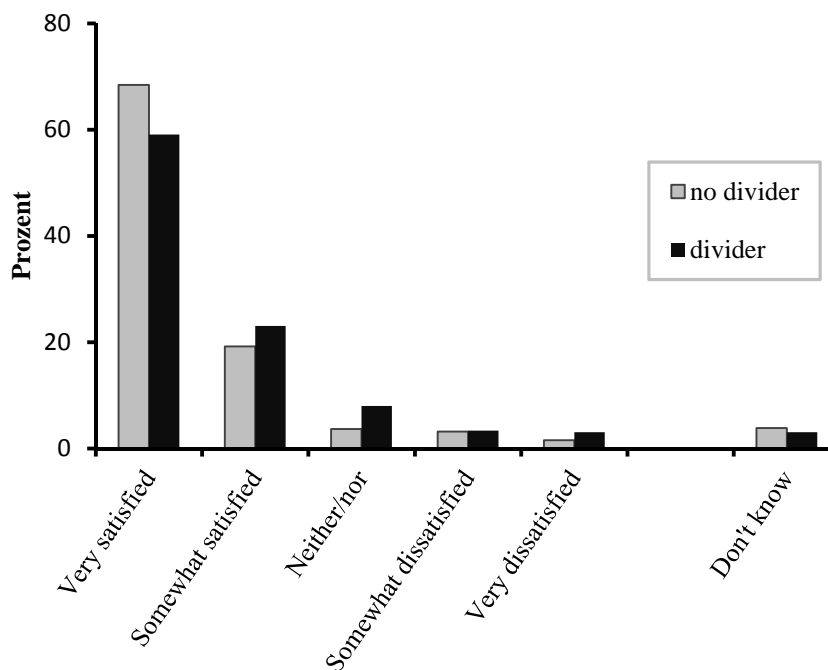
5. *Antwortverteilungen zu den Items*

Die Antwortverteilung in der Studie von Tourangeau et al. (2004) ist - unabhängig vom Skalenlayout - linksschief. Die Mehrheit der Befragten wählte eine Antwortkategorie auf der rechten Skalenseite, wo auch die graphische Manipulation der Skala stattfindet (vgl. Abbildung 3). In der Studie von Christian et al. (2009) sind die Verteilungen rechtsschief. Vor allem zu Item 2 antwortete die überwiegende Mehrheit der Befragten anhand der ersten Antwortkategorie (vgl. Abbildung 4).



Quelle: Tourangeau et al. (2004)

Abbildung 3 Linksschiefe Antwortverteilung des Item 1 bei Tourangeau et al. (2004)



Quelle: eigene Darstellung, basierend auf den publizierten Daten von Christian et al. (2009)

Abbildung 4 Rechtsschiefe Antwortverteilung des Item 2

In ihrer Untersuchung der Blickbewegungen von Befragten bei der Fragebeantwortung anhand von fünfstufigen, vollverbalisierten Ratingskalen zeigen Menold et al. (2014), dass ein Großteil der Befragten lediglich die Skalenlabels vom linken Ende bis zur Mitte der Skala fixiert. Die Labels am rechten Ende der Skala hingegen werden nicht betrachtet, und wenn doch, dann deutlich kürzer als die Labels am linken Skalenende. Dieser Effekt kann dadurch zustande kommen, dass

die meisten Befragten eine Antwortkategorie links von der Skalenmitte wählen, was sich in rechtsschiefen Verteilungen der untersuchten Items zeigt. Auch eine Studie der Blickbewegungen von Galesic et al. (2008a) kommt zu dem Ergebnis, dass Befragte sowohl in Antwortlisten als auch in Ratingskalen häufig lediglich die oberen bzw. linken Antwortalternativen betrachten. Aus diesen Ergebnissen kann geschlossen werden, dass Befragte eine Antwortskala nicht in ihrer vollen Breite wahrnehmen und somit auch nicht die DK-Kategorie, wenn sie ihre Antwort bereits unter den ersten Antwortkategorien ausmachen. Diese Annahme wird auch durch die Ergebnisse von Stern et al. (2007) unterstützt, die besagen, dass die DK-Kategorie häufiger gewählt wird, wenn die Mehrheit der Antworten anhand der letzten beiden Kategorien am Skalenende gegeben wurden, die Antwortverteilung also linksschief ist. Sie schreiben diesen Effekt der Tatsache zu, dass in dieser Situation die DK-Kategorie für Befragte visuell zugänglicher ist, als wenn die Antworten anhand von Kategorien am Skalenanfang gegeben werden.

Dies kann eine Erklärung dafür sein, dass Christian et al. (2009) keinen Effekt der graphischen Darstellung der DK-Kategorie in ihrem Experiment beobachten. Die meisten Befragten antworten direkt anhand der ersten oder zweiten Antwortkategorie und nehmen somit wahrscheinlich weder den verschobenen visuellen Mittelpunkt noch die graphisch betonte DK-Kategorie richtig wahr. In der Studie von Tourangeau et al. (2004) ist die Situation genau umgekehrt, die meisten Befragten wählen eine der letzten beiden Antwortkategorien, passieren somit den visuellen Mittelpunkt und nehmen die DK-Kategorie wahr.

5. Forschungsfragen & Hypothesen

Wie im vorangehenden vierten Abschnitt ausführlich beschrieben ist, kommen die Studien von Tourangeau et al. (2004) und Christian et al. (2009) zu widersprüchlichen Ergebnissen hinsichtlich des Einflusses der graphischen Präsentation der DK-Kategorie in Ratingskalen auf das Antwortverhalten von Befragten. Darüber hinaus bestehen keine hinreichenden Kenntnisse über die individuellen und situationellen Determinanten, welche Befragte mehr oder weniger

empfänglich für nonverbale, graphische Skalenelemente werden lassen. Diese Forschungslücke zu verringern, ist Ziel dieser Dissertation.

In dieser Dissertation werden in zwei empirischen Studien die Effekte des Skalenlayouts auf das Antwortverhalten von Befragten experimentell untersucht. Zudem werden als Determinanten von Layouteffekten in Studie 1 die Einstellungsstärke gegenüber den Frageinhalten und in Studie 2 Alter und Bildung betrachtet. Im Folgenden werden die jeweiligen Forschungsfragen und Designs der beiden Studien behandelt. Eine Beschreibung der verwendeten Daten und Methoden der jeweiligen Studien erfolgt im späteren Teil der Arbeit für Studie 1 in Abschnitt (6) und für Studie 2 in Abschnitt (12).

5.1 Studie 1

Studie 1 basiert auf den Daten eines Experiments, bei welchem Studenten der Politikwissenschaft der Justus-Liebig-Universität Gießen sowie der Universität Mannheim die Papierfragebögen selbständig ausfüllten. Die Mit der Studie 1 werden zwei Ziele verfolgt:

Erstens werden die Effekte der graphischen Präsentation der DK-Kategorie einer fünfstufigen Ratingskala auf das Antwortverhalten der Befragten ausführlich untersucht. Dabei bildet das Antwortverhalten die abhängige Variable der Untersuchung und wird als Tendenz zur Wahl der rechten Extrem-, Mittel- und DK-Kategorie operationalisiert. Da es sich um eine fünfstufige Ratingskala mit DK-Kategorie handelt, könnte auch das Antwortverhalten zu allen fünf Antwortkategorien und der DK-Kategorie betrachtet werden. Allerdings werden die stärksten Effekte des Skalenlayouts auf die Wahl der Extremkategorien am rechten Skalenrand, auf die Mittel- und auf die DK-Kategorie erwartet, wie es in den Hypothesen in Unterabschnitt (5.3) erläutert wird. Zudem würde eine Betrachtung aller fünf Antwortkategorien und der DK-Kategorie den Umfang dieser Dissertation überschreiten. Deshalb erfolgt eine Einschränkung der Untersuchung des Antwortverhaltens auf die drei genannten Antwortkategorien.

Zweitens wird in Studie 1 betrachtet, inwieweit die individuelle Einstellungsstärke von Befragten diese gegenüber den Frageinhalten der untersuchten Items mehr oder

weniger empfänglich für die nonverbalen, graphischen Elemente aus der Darstellung der DK-Kategorie werden lässt.

Die Variationen der graphischen Darstellung der DK-Kategorie (vgl. Abbildung 5) gleichen den Layouts der beiden Studien von Tourangeau et al. (2004) und Christian et al. (2009) dahingehend, dass die DK-Kategorie entweder mittels eines Trennstrichs von der Ratingskala abgesetzt (SEP-DK-Layout) oder einfach an die Ratingskala angehängt wird (ADD-DK-Layout). Darüber hinaus wird in Studie 1 noch ein drittes Layout, eine Ratingskala ohne DK-Kategorie, aufgenommen (NO-DK-Layout). Damit ist der Vergleich des Antwortverhaltens von Befragten zu Skalen mit und ohne DK-Kategorie möglich.

NO-DK-Layout		Keine DK-Kategorie			
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
SEP-DK-Layout		Abtrennung mittels eines Trennstrichs			
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	Kann ich nicht sagen
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ADD-DK-Layout		Keine graphische Abtrennung			
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	Kann ich nicht sagen
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 5 Experimentell variierte Skalenlayouts der Studie 1 und Studie 2

Ein weiterer Unterschied im Aufbau des Experiments der Studie 1 gegenüber den Studien von Tourangeau et al. (2004) und Christian et al. (2009) ist, dass die Effekte des Skalenlayouts auf das Antwortverhalten nicht anhand einzelner Items untersucht werden. Vielmehr werden zwei Mehr-Item-Skalen herangezogen, die in den Sozialwissenschaften gängig verwendet werden: Die erste Skala besteht aus sieben Items, welche das Konstrukt der Politischen Selbstwirksamkeit, im Englischen „Political Efficacy“ (PES), erfassen (zum Konzept und zur Itementwicklung vgl. auch: Craig et al. 1990). Eine deutschsprachige Version der PES wurde

beispielsweise in der Allgemeinen Bevölkerungsbefragung der Sozialwissenschaften (ALLBUS) von 2008 erhoben. Die zweite Skala umfasst zehn Items, welche die beiden Subskalen „Zeitmanagement“ und „Aufmerksamkeit“ des Inventars zur Erfassung von Lernstrategien im Studium (LIST) repräsentieren. Das LIST-Inventar ist ein standardisierter Fragebogen, der von Wild und Schiefele (1994) entwickelt wurde, um kognitive, metakognitive und ressourcenbezogene Lernstrategien zu erfassen. Für jede Skala erfolgt die Darstellung der Items in einer Itembatterie.

Der Vorteil der Verwendung mehrerer Items, die gemeinsam ein latentes Konstrukt erfassen, gegenüber Experimenten an einzelnen Items besteht darin, dass weitere Einflussgrößen, wie Frageformulierung, eine starke rechts- oder linksschiefe Antwortverteilung oder fragespezifische Probleme des Befragten im Mittel kontrolliert werden. Bei der Auswahl der beiden Skalen waren ihre thematischen Inhalte leitend. Die PES-Skala umfasst verschiedene Aspekte der Politischen Selbstwirksamkeit, für die erwartet wird, dass die meisten Befragten keine klar vorliegende Einstellung dazu besitzen. Dieselbe Annahme gilt auch für die Items über die Frauenrechte, die in der Studie von Tourangeau et al. (2004) Verwendung finden (vgl. Unterabschnitt 4.1). Christian et al. (2009) hingegen verwenden ein Item, das die Zufriedenheit mit dem kulturellen Angebot auf dem Campus abfragt (vgl. Unterabschnitt 4.2). Ein Thema, welches für die meisten Studierenden zum Alltag gehören sollte. Daher wird als zweite Skala in Studie 1 die LIST-Skala gewählt, deren Items Lernstrategien beschreiben, die ebenfalls zum Studienalltag von Studenten gehören.

Auch die Auswertung der Experimentaldaten ist in dieser Studie eine andere. Tourangeau et al. (2004) und Christian et al. (2009) basieren ihre Auswertung im Wesentlichen auf Mittelwertvergleiche und Vergleiche der Item Nonresponse-Rate zwischen den beiden Layoutgruppen. In der Auswertung der Daten des PAPI-Experiments der Studie 1 hingegen werden die Tendenzen der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie, für die Unterschiede in den Hypothesen in Unterabschnitt (5.3) formuliert werden, anhand von Multilevel-Modellen über alle untersuchten Items hinweg betrachtet. Somit stehen für jeden Befragten Daten über das Antwortverhalten über siebzehn Items hinweg zur Verfügung, welche dann

unter Berücksichtigung des Ratingskalenlayouts und weiterer Variablen auf Befragten- und Itemebene ausgewertet werden können (vgl. Unterabschnitt 6.3.2).

Schließlich ist die Studie 1 dieser Dissertation die erste Studie über Layouteffekte im deutschsprachigen Raum.

5.2 Studie 2

Studie 2 basiert auf den Daten eines Experiments, welches im Rahmen der elften Erhebung des GESIS Online Access Panel Pilots (vgl. auch: Struminskaya et al. 2014) erhoben wurde. Das Hauptziel der Studie 2 ist zu untersuchen, ob die in Studie 1 anhand des PAPI-Experiments mit Studenten gefundenen Layouteffekte in einer heterogenen Erwachsenenstichprobe im Online-Modus repliziert werden. Daher werden auch wieder die gleichen drei Skalenlayouts (vgl. Abbildung 5) wie in Studie 1 experimentell variiert. Ein zweites Ziel ist, den Einfluss von Alter und Bildung auf das Auftreten von Layouteffekten zu analysieren.

Das Experiment der Studie 2 wird an zehn Items durchgeführt, welche die Einstellungen gegenüber der Europäischen Union erheben (EU-Items). Diese entstammen der German Longitudinal Election Studie 2011. Im Unterschied zur Studie 1, in der die Items einer Skala gemeinsam in einer Antwortmatrix dargestellt werden, wird in diesem Experiment pro Webseite nur ein einzelnes Item mit der entsprechenden Antwortskala abgebildet (sogenannte Single-Item-Darstellung).

Auch in Studie 2 wird die Tendenz zur Wahl der rechten Extremen-, der Mittel- und der DK-Kategorie über alle zehn experimentell variierten Items anhand von logistischen Multilevel-Modellen untersucht (vgl. Unterabschnitt 12.3).

5.3 Hypothesen

1. *Hypothesen, die sowohl in Studie 1 als auch in Studie 2 untersucht werden:*

Effekte des Skalenlayouts auf das Antwortverhalten

Die Heuristik „middle means typical“ umfasst die Annahme, dass Befragte die Mitte einer Skala als den typischen Wert der Population betrachten und sich bei der Verwendung von Ratingskalen eher am visuellen als am konzeptionellen Mittelpunkt orientieren (vgl. Unterabschnitt 1.3), vgl. auch: Tourangeau et al. 2004). Bezüglich dieser Heuristik werden folgende drei Hypothesen über die Einflussnahme des Skalenlayouts auf das Antwortverhalten der Befragten in Studie 1 und Studie 2 verfolgt:

H 1.1 – Wahl der rechten Extremkategorie

Für das ADD-DK-Layout wird aufgrund der Verschiebung des visuellen Mittelpunktes in Richtung des rechten Skalenendes eine Verschiebung der Antworten - ebenfalls hin zum rechten Skalenende – erwartet, was sich in einer höheren Rate an Antworten zu der rechten Extremkategorie niederschlagen sollte.

Daher lautet die Hypothese H 1.1, dass im ADD-DK-Layout die Extremkategorie am rechten Ende der Skala häufiger gewählt wird im Vergleich zum NO-DK- und SEP-DK-Layout.

Zudem wird erwartet, dass dieser Effekt auch dadurch verstärkt wird, dass Befragte die Tendenz besitzen, die äußere Extremkategorie zu vermeiden (Kerlinger 1973; Tourangeau et al. 2000). Im ADD-DK-Layout liegt die Extremkategorie nicht am äußeren, rechten Skalenende, da auf sie noch die DK-Kategorie folgt. Es wird angenommen, dass die rechte, extreme Antwortkategorie daher dem Befragten visuell als nicht so extrem erscheint und deshalb in diesem Layout häufiger im Vergleich zu den anderen beiden Layouts gewählt wird.

H 1.2 – Wahl der Mittelkategorie

Im NO-DK- und SEP-DK-Layout stimmen der visuelle Skalenmittelpunkt und die mittlere Antwortkategorie miteinander überein (vgl. Abbildung 5, Seite 38). Befragte, die eine mittlere Antwort geben möchten – sei es aus inhaltlichen Gründen oder als Ausdruck von Satisficing – finden die entsprechende Antwortkategorie in der visuellen Mitte der Skala vor. Anders ist dies im ADD-DK-

Layout, wo die visuelle Skalenmitte und die mittlere Antwortkategorie nicht miteinander übereinstimmen. Befragte in diesem Layout, die eine mittlere Antwort geben möchten, können die Heuristik „middle means typical“ nicht einfach anwenden, da hier keine visuelle Mitte vorliegt.

Die Hypothese 1.2 lautet, dass im ADD-DK-Layout im Vergleich zum NO-DK- und SEP-DK-Layout weniger häufig die mittlere Antwortkategorie gewählt wird.

H 1.3 – Wahl der DK-Kategorie

Es wird angenommen, dass durch die graphische Absetzung der DK-Kategorie im SEP-DK-Layout diese visuell betont wird und somit für Befragte, welche eine DK-Antwort geben möchten – aus inhaltlichen Gründen oder wegen Satisficing – deutlich wahrnehmbar ist (vgl. Abbildung 5, Seite 38). Ein anderer Effekt wird für das ADD-DK-Layout angenommen, wo die DK-Kategorie in ihrem Layout der inhaltlichen Ratingskala gleicht.

Daher besagt die Hypothese H 1.3, dass im SEP-DK-Layout die DK-Kategorie häufiger gewählt wird als im ADD-DK-Layout.

2. Hypothesen, die ausschließlich in Studie 1 untersucht werden

H 2 – Unterschiedliche Layouteffekte für PES- und LIST-Skala

Für die Politische Selbstwirksamkeit, erfasst durch die Items der PES-Skala, wird angenommen, dass sie ein abstraktes Konstrukt darstellt, zu dem die Befragten zumeist keine klar vorliegenden Einstellungen besitzen. Vielmehr ist die Annahme, dass die Einstellung zur Politischen Selbstwirksamkeit im kognitiven Fragebeantwortungsprozess generiert und dann anhand der angebotenen Antwortskala berichtet wird. Im Gegensatz dazu werden die Studenten bei der Befragung zu ihren Lernstrategien im Studium, hier erhoben durch die Items der LIST-Skala, mit Verhaltensfragen konfrontiert, die sich auf ihren Lebensalltag beziehen. Es ist anzunehmen, dass die Beantwortung dieser Fragen bei den meisten Befragten unmittelbar aus dem Gedächtnis erfolgen kann, ohne dass dazu noch tiefergehende kognitive Leistungen mit Unterstützung von temporär verfügbaren Informationen vollbracht werden müssen (vgl. Unterabschnitt 4.3).

Deshalb wird hier die Hypothese H 2 getroffen, dass Befragte bei Beantwortung der Items der PES empfänglicher für das Skalenlayout sind als bei Beantwortung der Items des LIST-Inventars.

H 3 – Individuelle Einstellungsstärke als Moderator von Layouteffekten

Befragte mit einer hohen Einstellungsstärke zeigen sich weniger empfänglich für Effekte verbaler und nonverbaler Fragebogenelemente (vgl. Unterabschnitt 2.2.1, vgl. auch: Schwarz 1996; Tourangeau et al. 2004,2007; Rockwood et al. 1997). Die Frage ist nun, ob mit sinkender Einstellungsstärke Befragte auch empfänglich für das Layout von Antwortskalen werden, ob also Einstellungsstärke ein Moderator für das Auftreten von Layouteffekten ist. Hierzu existiert bislang noch keine Untersuchung.

Aufgrund dieser getroffenen Annahmen lautet die Hypothese H 3, dass mit zunehmender Einstellungsstärke Befragte weniger das Ratingskalenlayout bei dem kognitiven Fragebeantwortungsprozess einbeziehen und daher mit zunehmender Einstellungsstärke weniger bis gar keine Layouteffekte auftreten.

3. Hypothesen, die ausschließlich in Studie 2 untersucht werden

Kognitive Leistungsfähigkeit als Moderator von Layouteffekten

Verschiedene Studien zeigen, dass Befragte mit einer geringeren kognitiven Leistungsfähigkeit eher Probleme mit dem Verständnis von Fragen oder Antwortskalen haben als Befragte mit einer höheren kognitiven Leistungsfähigkeit (vgl. Unterabschnitt 2.2.1). Zur Unterstützung des Frageverständnisses ziehen solche Befragte weitere Informationen aus dem Umfeld der Frage heran, beispielsweise Informationen aus Inhalten vorhergehender Fragen oder aus verbalen oder numerischen Etiketten der Antwortskala (vgl. u.a. Schwarz et al. 1991).

Die Annahme lautet daher, dass Befragte mit einer geringeren kognitiven Leistungsfähigkeit das Layout von Antwortskalen stärker als Informationsquelle zur Unterstützung bei der Fragebeantwortung nutzen als Befragte mit einer höheren kognitiven Leistungsfähigkeit.

Indikatoren für eine starke bzw. schwache kognitive Leistungsfähigkeit sind in Studie 2 „Bildung“ und „Alter“ der Befragten.

H 4.1 – Bildung

Menschen mit einer höheren Bildung besitzen in der Regel eine höhere kognitive Leistungsfähigkeit als Menschen mit einer niedrigen Bildung (Asendorpf und Neyer 2012).

Daher lautet die Hypothese H 4.1, dass Befragte mit einem höheren Bildungsabschluss weniger empfänglich für das Skalenlayout sind als Befragte mit einem niedrigen Bildungsabschluss (vgl. Unterabschnitt 2.2.1).

H 4.2 – Alter

Aus der Kognitionsforschung ist bekannt, dass die kognitive Leistungsfähigkeit mit zunehmendem Alter abnimmt (Salthouse 1991,2004).

In Bezug dazu wird die Hypothese H 4.2 formuliert, dass für jüngere Befragte weniger Layouteffekte zu beobachten sind als für ältere Befragte (vgl. Unterabschnitt 2.2.1).

III Studie 1

Layouteffekte in einer papierbasierten Befragung von Studenten

6. Studiendesign

In Studie 1 werden die Effekte der graphischen Darstellung der DK-Kategorie in einer fünfstufigen Ratingskala in Papierfragebögen untersucht, die von Studenten selbständig ausgefüllt wurden. Des Weiteren ist es Ziel dieser Studie 1, Kenntnisse über den Einfluss des Item-Themas sowie der individuellen Stärke der Einstellung gegenüber dem abgefragten Thema auf das Auftreten von Layouteffekten zu erlangen.

In dem Experiment werden drei Layouts einer fünfstufigen Ratingskala experimentell variiert. Im ersten wird keine DK-Kategorie angeboten (NO-DK-Layout), im zweiten wird die DK-Kategorie graphisch mittels eines Trennstriches von den inhaltlichen Kategorien abgetrennt dargestellt (SEP-DK-Layout) und im dritten Layout wird die DK-Kategorie schließlich ohne graphische Abtrennung von den inhaltlichen Antwortkategorien angeboten (ADD-DK-Layout) (vgl.

Abbildung 5, Seite 38). Die beiden zuletzt genannten Layouts sind somit eine Replikation der in den Studien von Tourangeau et al. (2004) und Christian et al. (2009) verwendeten Designs. Die Unterschiede zwischen Studie 1 dieser Dissertation und den beiden bereits publizierten Studien sind in Unterabschnitt (5.1) in Kapitel II erläutert.

Der folgende Unterabschnitt (6.1) beschreibt das Verfahren der Datengewinnung und die Stichprobenstatistik. Unterabschnitt (6.2) stellt die Items vor, anhand welcher das Experiment durchgeführt wurde und welche die abhängigen und

unabhängigen Variablen der späteren Analyse bilden. Die Methode der Datenanalyse wird schließlich im letzten Unterabschnitt (6.3) erläutert.

6.1 Daten

Das Experiment war in eine größere Studie zu Einstellungen gegenüber politischen und sozialen Themen sowie zur Persönlichkeit eingebettet. Zu Beginn des Jahres 2012 wurden Studenten der Universität Gießen und der Universität Mannheim im Rahmen von Vorlesungen der jeweiligen politikwissenschaftlichen Fakultät mittels Papierfragebögen befragt. Insgesamt haben 495 Studenten die Papierfragebögen selbstständig ausgefüllt. Aufgrund eines parallel laufenden Experiments haben 307 der 495 Studenten Fragebögen mit einem der drei hier behandelten Skalenlayouts erhalten. Die Layouts wurden den Studenten zufällig zugeteilt.

Die folgende Stichproben-Statistik bezieht sich demnach auf die Angaben dieser 307 Befragten. 51,5 % der Befragten sind männlich, 47 % weiblich und 1,5 % machten keine Angabe bezüglich ihres Geschlechts. Das durchschnittliche Alter der Befragten beträgt 23 Jahre ($SD = 5,77$). Die Befragten in den drei experimentellen Layoutgruppen unterscheiden sich nicht signifikant hinsichtlich ihres Geschlechts ($\chi^2=0,89$, $p>0,10$), Alters ($\chi^2=42,89$, $p>0,10$), oder ihrer Studiendisziplin ($\chi^2=2,77$, $p>0,10$). Die drei Antwortskalenlayouts (NO-DK, SEP-DK und ADD-DK) wurden zufällig an die Studenten verteilt; 35 % komplettierten einen Fragebogen im NO-DK-Layout ($N = 107$), 33,5 % im SEP-DK-Layout ($N = 103$), und 31,5 % im ADD-DK-Layout ($N = 97$).

6.2 Items

Insgesamt wurden in Studie 1 der Dissertation die Ratingskalen von siebzehn Items, die zwei in der sozialwissenschaftlichen Forschung gängigen Skalen entstammen, experimentell variiert. Die erste Skala erfasst dabei die politische Selbstwirksamkeit, die im Englischen als „Political Efficacy Scale“ bezeichnet und hier PES abgekürzt wird, und besteht aus sieben Items, die dem ALLBUS 2008 entnommen wurden. Die zweite Skala erhebt anhand von vier Items die Subskala „Zeitmanagement“ und anhand von sechs Items die Subskala „Aufmerksamkeit“

des Inventars zur Messung von Lernstrategien im Studium (kurz LIST). Das LIST-Inventar ist ein standardisierter Fragebogen, der von Wild und Schiefele (Wild und Schiefele 1994) entwickelt wurde, um kognitive, metakognitive und ressourcenbezogene Lernstrategien zu erfassen.

Sowohl die PES- als auch die LIST-Items wurden nicht im Einzel-Item-Format sondern in einer Antwortmatrix dargestellt.

Die sieben Items der PES sind:

1. Die Politiker kümmern sich nicht viel darum, was Leute wie ich denken.
2. Ich traue mir zu, in einer Gruppe, die sich mit politischen Fragen befasst, eine aktive Rolle zu übernehmen.
3. Die ganze Politik ist so kompliziert, dass jemand wie ich gar nicht versteht, was vorgeht.
4. Die Politiker bemühen sich im Allgemeinen darum, die Interessen der Bevölkerung zu vertreten.
5. Im Allgemeinen weiß ich eher wenig über Politik.
6. Die meisten Leute wären durchaus fähig, in einer politischen Gruppe mitzuarbeiten.
7. Die Durchschnittsbürger verstehen nur wenig von Politik.

Die Antwortkategorien der fünfstufigen, vollverbalisierten Ratingskala sind „lehne stark ab“, „lehne eher ab“, „weder noch“, „stimme eher zu“ und „stimme stark zu“. Die DK-Kategorie ist mit „kann ich nicht sagen“ bezeichnet.

Die zehn Items der LIST-Skala sind:

Subskala „Zeitmanagement“

1. Beim Lernen halte ich mich an einen bestimmten Zeitplan.
2. Ich lege bestimmte Zeiten fest, zu denen ich dann lerne.
3. Ich lege die Stunden, die ich täglich mit Lernen verbringe, durch einen Zeitplan fest.
4. Ich lege vor jeder Lernphase eine bestimmte Zeitdauer fest.

Subskala „Aufmerksamkeit“

5. Beim Lernen merke ich, dass meine Gedanken abschweifen.

6. Es fällt mir schwer, bei der Sache zu bleiben.
7. Ich ertappe mich dabei, dass ich mit meinen Gedanken ganz woanders bin.
8. Beim Lernen bin ich unkonzentriert.
9. Wenn ich lerne, bin ich leicht abzulenken.
10. Meine Konzentration hält nicht lange an.

Die Antwortkategorien der fünfstufigen, vollverbalisierten Ratingskala sind „trifft überhaupt nicht zu“, „trifft wenig zu“, „trifft teilweise zu“, „trifft ziemlich zu“ und „trifft voll und ganz zu“. Die DK-Kategorie ist mit „kann ich nicht sagen“ bezeichnet.

Die Stärke der Einstellung gegenüber der politischen Selbstwirksamkeit und gegenüber den persönlichen Lernstrategien kann anhand verschiedener Attribute gemessen werden, die inhaltlich klar voneinander unterscheidbar sind und verschiedene Aspekte der Einstellung ausdrücken (für einen Überblick vgl. Krosnick und Smith 1994). In dem Experiment der Studie 1 werden die zwei Attribute „Sicherheit“ und „Wichtigkeit“ erfasst, die das subjektive Befinden des Befragten hinsichtlich der abgefragten Einstellung beschreiben. Das Attribut „Sicherheit“ bezieht sich dabei auf die Überzeugung eines Befragten, dass seine Einstellung gegenüber dem abgefragten Objekt korrekt ist. Das Attribut „Wichtigkeit“ bezieht sich darauf, wie wichtig eine Person ihre Einstellung bzw. das Thema für sich persönlich empfindet (Krosnick und Smith 1994). Die Items zu „Sicherheit“ und „Wichtigkeit“ wurden, wie in der Studie von Bassili und Krosnick (2000), jeweils direkt nach den sieben PES-Items bzw. direkt nach den zehn LIST-Items gestellt.

Das Item über die Sicherheit bei der Antwortgabe lautet:

PES: Wie sicher waren Sie sich bei der Beantwortung der Fragen zur Einschätzung der politischen Einflussnahme?

LIST: Wie sicher waren Sie sich bei der Beantwortung der Fragen über Ihr Lernen?

Die Kategorien der fünfstufigen, vollverbalisierten Antwortskala sind „überhaupt nicht sicher“, „wenig sicher“, „mittelmäßig sicher“, „ziemlich sicher“ und „sehr sicher“.

Das Item über die persönliche Wichtigkeit des Themas lautet:

PES: Wie wichtig ist für Sie die Möglichkeit zur Einflussnahme auf politische Entscheidungen?

LIST: Wie wichtig ist für Sie persönlich zu wissen, wie Sie lernen?

Die Antwortkategorien der fünfstufigen, vollverbalisierten Antwortskala sind „überhaupt nicht wichtig“, „wenig wichtig“, „mittelmäßig wichtig“, „ziemlich wichtig“ und „sehr wichtig“.

Die beiden Items werden aggregiert zu einem multiplikativen Index, bei dem für jeden Fall der Wert der Sicherheit mit dem der Wichtigkeit multipliziert wird. Daher weist der Index der Einstellungsstärke einen theoretischen Wertebereich von 1 (sehr schwache Einstellung) bis 25 (sehr starke Einstellung) auf. Es wird ein multiplikativer Index gewählt, weil mit diesem Verfahren der Indexbildung gegeben ist, dass die Abwesenheit eines der beiden Attribute den Minimalwert bedingt. Bei der additiven Indexbildung ist dies hingegen nicht der Fall, hier kann ein Befragter auch bei Abwesenheit eines der beiden Attribute noch einen Indexwert im mittleren Bereich erzielen (Schnell et al. 2011). Ein multiplikativer Index gibt eine angemessene inhaltliche Beschreibung des Konzepts der Einstellungsstärke wieder, denn eine Einstellung kann nicht als stark bezeichnet werden, wenn ein Befragter zwar hinsichtlich seiner Antwort sicher, das Thema selbst ihm aber nicht wichtig ist. Umgekehrt ist die Stärke der Einstellung ebenfalls als gering zu bewerten, wenn ein Befragter bei einem ihm wichtigen Thema Unsicherheit bei der Antwortgabe verspürt. Diese Beziehung (Interaktion) zwischen Sicherheit und Wichtigkeit kann also nur durch eine Multiplikation beider Attributwerte korrekt dargestellt werden, denn somit können geringe Werte eines Attributs nicht durch hohe Werte des anderen Attributs kompensiert werden. Solche multiplikativen Indizes finden auch Verwendung in der Motivationsforschung, wo die Werte von Variablen bezüglich der Evaluation von Themen mit der Wichtigkeit des Themas für die persönliche Intention und das persönliche Verhalten multipliziert werden (vgl. u.a. Trautwein et al. 2012; Vroom 1964).

6.3 Methode

Studie 1 untersucht den Einfluss des Ratingskalenlayouts auf das Antwortverhalten der befragten Personen unter Einbeziehung weiterer erklärender Variablen. Dabei wird das Antwortverhalten, welches operationalisiert ist, als Wahl der rechten Extrem-, der Mittel- bzw. der DK-Kategorie über die sieben Items der PES und über die zehn Item des LIST-Inventars hinweg beobachtet. Demnach handelt es sich um siebzehn (wiederholte) Messungen des Antwortverhaltens für jeden Befragten, weshalb die Daten eine hierarchische Struktur aufweisen. Für die Analyse der Daten können verschiedene Methoden angewandt werden, was im folgenden Unterabschnitt (6.3.1) kurz erörtert wird. Zudem wird hierbei auch die Auswahl der Analysemethode „Multilevel-Regression“ begründet. In Unterabschnitt (6.3.2) ist erläutert, wie das Verfahren der logistischen Multilevel-Regression in Bezug zu den Daten und den Zielen der Studie 1 steht. Unterabschnitt (6.3.3) stellt die sogenannten „marginalen Effekte“ vor, welche eine von verschiedenen Kenngrößen logistischer Regressionsmodelle darstellen und anhand derer die Analyseergebnisse dieser Studie interpretiert werden. Im letzten Unterabschnitt (6.3.4) wird ein Überblick über den Aufbau der logistischen Multilevel-Regressionsmodelle der Studie 1 geboten.

6.3.1 Auswahl der Analysemethode

Um die Hypothesen über die Unterschiede der Wahl der rechten extremen, der mittleren und der DK-Kategorie in Abhängigkeit des Antwortskalenlayouts und weiterer Kovariate zu testen, bieten sich hier zwei alternative multivariate Analyseverfahren an. Das erste Verfahren ist die Poisson-Regression, welche zur Analyse von Zählvariablen als abhängige Variable eingesetzt wird (vgl. u.a. Cameron und Trivedi 1998; Tutz 2010). Bei diesem Verfahren wird zur Bildung der abhängigen Variable über alle siebzehn experimentell variierten Items eine Zählvariable erstellt. Diese gibt an, wie häufig ein Befragter die interessierende Antwortkategorie gewählt hat. Demnach besitzt die abhängige Variable einen theoretischen Wertebereich von Null (die Kategorie wurde für keines der Items gewählt) bis siebzehn (die Kategorie wurde für alle Items gewählt). Durch dieses Vorgehen wird die Information über die Antwortgabe zu den Items von der Item-

auf die Befragtenebene aggregiert. Wird dieses Analyseverfahren angewandt, muss allerdings ein starker Informationsverlust über die individuelle Heterogenität in Kauf genommen werden, der bei der Aggregation der Information entsteht.

Das zweite Analyseverfahren ist die Multilevel-Regression, welches es vermag, die hierarchische Struktur der Daten zu berücksichtigen und den Einfluss von Faktoren auf verschiedenen Ebenen der hierarchischen Struktur auf die abhängige Variable zu betrachten (vgl. u.a. Hox 2010; Rabe-Hesketh und Skrondal 2008). In Studie 1 bilden die siebzehn untersuchten Items den Level - 1 und sind unter den Befragten auf Level - 2 gruppiert (vgl. Abbildung 6). Somit kann auf Befragtenebene über alle variierten Items die Neigung untersucht werden, die rechte extreme, die mittlere bzw. die DK-Kategorie zu wählen. Gleichzeitig können auch Einflussgrößen auf der Ebene der Items berücksichtigt werden, wie z.B. eine Zuteilung der Items zu Konstrukten (PES / LIST) und die Einstellungsstärke gegenüber der politischen Selbstwirksamkeit und gegenüber den persönlichen Lernstrategien (vgl. Abbildung 6). Des Weiteren können auch Interaktionen der Faktoren auf Item- und Befragtenebene, wie zum Beispiel Layout \times Einstellungsstärke, betrachtet werden.

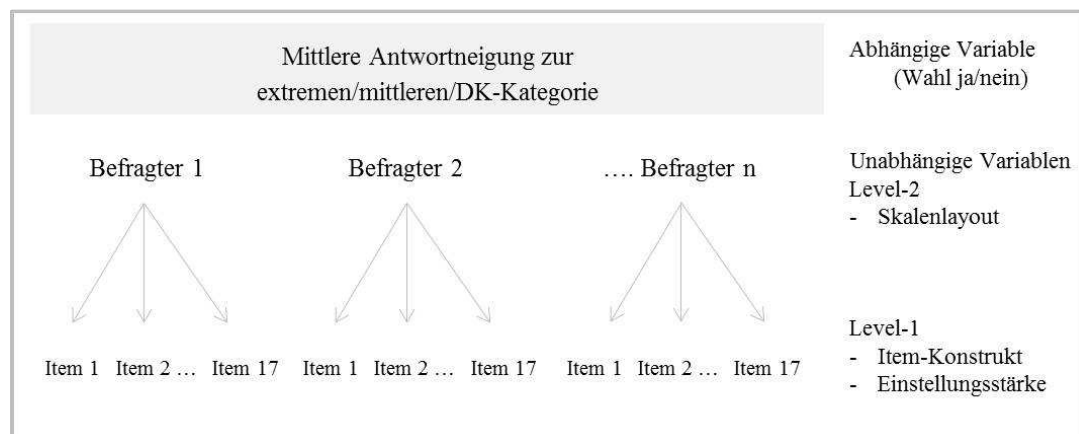


Abbildung 6 Multilevel-Modell der Gruppierung von experimentell variierten Items in Befragten und den Level-1 und Level-2 Variablen der Studie 1

Da die Multilevel-Regression gegenüber der Poisson-Regression mehr der in den Daten enthaltenen Informationen verwertet und zudem die Berücksichtigung von Erklärungsfaktoren sowohl auf Item- als auch auf Befragtenebene ermöglicht, wird sie zur Analyse der Daten der Studie 1 dieser Dissertation eingesetzt.

6.3.2 Logistische Multilevel-Regression mit fixen Effekten

Die drei abhängigen Variablen der Studie 1 bringen zum Ausdruck, ob ein Befragter die interessierenden Antwortkategorien „rechte Extremkategorie“, „Mittelkategorie“ bzw. „DK-Kategorie“ zur Antwortabgabe gewählt hat oder nicht gewählt hat. Es handelt sich demnach um dichotome abhängige Variablen, für deren Analyse sich das Verfahren der logistischen Regression eignet (vgl. u.a. Backhaus et al. 2006; Best und Wolf 2010; Hosmer und Lemeshow 2000). In logistischen Modellen werden die Regressions-Koeffizienten mittels des Maximum-Likelihood-Verfahrens geschätzt. Dieses Verfahren zielt auf eine Schätzung ab, die die Wahrscheinlichkeit maximiert, eine möglichst gute Übereinstimmung zwischen der Verteilung der vorhergesagten Werte der abhängigen Variable mit der beobachteten Verteilung der Werte zu erlangen (vgl. u.a. Eliason 1993; Gautschi 2010). Es resultiert schließlich in der Ausgabe von Logit-Koeffizienten, welche den Zusammenhang zwischen der abhängigen und den unabhängigen Variablen beschreiben und Informationen über die statistische Signifikanz der Zusammenhänge geben. Für Studie 1 zeigen beispielsweise die Logit-Koeffizienten für die verschiedenen Ausprägungen der nominalen Variable „Skalenlayout“ an, ob sie einen signifikanten Einfluss auf die mittlere Neigung der Wahl der interessierenden Antwortkategorien im Vergleich zur Referenzkategorie nehmen. Des Weiteren können in logistischen Regressionen, wie es auch in linearen OLS-Regressionen möglich ist, explizite Interaktionsterme als unabhängige Variable in das Modell aufgenommen werden. In Studie 1 sind die beiden Interlevel-Interaktionen „Skalenlayout \times Einstellungsstärke“ und „Skalenlayout \times Item-Konstrukt“ von Interesse. Dabei wird anhand der ersten Interaktion untersucht, ob der Effekt des Skalenlayouts auf die Wahrscheinlichkeit der Wahl der Antwortkategorie durch die individuelle Einstellungsstärke moderiert wird (ob es also bestimmte Bereiche der Einstellungsstärke gibt, bei denen Layouteffekte signifikant auftreten bzw. nicht auftreten). Die zweite Interaktion testet, ob Layouteffekte in den beiden betrachteten Item-Konstrukten PES und LIST gleichermaßen zu beobachten sind.³

³ Des Weiteren werden Dummy-Variablen für die experimentell variierten Items in das logistische Regressionsmodell aufgenommen. Denn aufgrund der wiederholten Durchführung der experimentellen Variation anhand mehrerer Items können die Ergebnisse von Dynamiken des Beantwortungsprozesses überlagert sein, die sowohl auf die unabhängigen Variablen als auch auf die abhängige Variable wirken. Unter Einbeziehung der Item-Dummies werden dann mögliche

Eine weitere Spezifikation des Analysemodells der Studie 1 ist darin begründet, dass für die Experimentaldaten angenommen werden muss, dass das Antwortverhalten bei einzelnen Items nicht vollständig durch die hier betrachteten Variablen modelliert wird. Dieser unberücksichtigte Fehler sollte nicht unabhängig über die betrachteten Items sein, d.h. das Antwortverhalten einer Person bei einem bestimmten Item sollte mit dem Antwortverhalten bei den nachfolgenden Items korreliert sein. Dieses Vorliegen von Korrelationen der Fehlerterme auf Level - 1 schließt die Verwendung von Random-effects-Modellen aus. Die siebzehn Items wurden sowohl zeitlich (ein gemeinsamer Erhebungszeitpunkt) als auch räumlich (je Konstrukt eine Item-Batterie) aufeinanderfolgend erhoben, was eine Korrelation der Fehlerterme mit sich bringt. Denkbar ist zum Beispiel, dass ein Befragter eine Antworttendenz zeigt, die dann die Beantwortung aller siebzehn Items systematisch beeinflusst. Auch Fragereihenfolgeeffekte bei der Beantwortung der direkt aufeinander folgenden Items können nicht ausgeschlossen werden. Daher wird in dieser Studie das Fixed-effects-Modell zur Datenanalyse herangezogen.⁴⁵

Die logistische Multilevel-Regression mit ausschließlich fixen Effekten entspricht der gepoolten multivariaten logistischen Regression (Snijders und Bosker 2012). Der Unterschied besteht lediglich darin, dass die Erklärungsfaktoren den verschiedenen Leveln zugeordnet werden können. Diese Struktur wird durch die Verwendung eines Datensatzes im sogenannten „long format“ repräsentiert, in dem die Werte der siebzehn experimentell variierten Items als Fälle betrachtet werden.

Unterschiede in den Achsenabschnitten der einzelnen Items kontrolliert. Dieser Effekt ist vor allem aus der Analyse von Paneldaten bekannt. Hier wird mittels Perioden-Dummies der allgemeine Trend über die Zeit kontrolliert, daher wird der Effekt hier als Periodeneffekt bezeichnet (Wooldrige 2002)

⁴ Multilevel-Regressionen können mit zufälligen Effekten (sogenannte Random-effects-Modelle) oder mit fixen Effekten (sogenannte Fixed-effects-Modelle) geschätzt werden, wobei auch Multilevel-Modelle möglich sind, die sowohl fixe als auch zufällige Effekte beinhalten (vgl. u.a. Hox 2010; Snijders und Bosker 2012). Random-effects-Modelle bieten zwar effizientere Schätzer als Fixed-effects-Modelle, aber nur unter Erfüllung der Annahme, dass keine Korrelation der Fehlerterme innerhalb von Level - 2 auf die Einheiten auf Level - 1 vorliegt. Wird diese Annahme nicht erfüllt, wie es in den beiden Studien der Dissertation der Fall ist, sind die Schätzer des Random-effects-Modells nicht mehr konsistent. In Fixed-effects-Modellen hingegen sind die Schätzer robust gegenüber einer solchen Autokorrelation. Mit anderen Worten wird die Effizienz der Random-effects-Modelle mit stärkeren Annahmen über die Fehlerterme erkauft.

⁵ Um die Korrelation der Fehlerterme zu berücksichtigen, werden in dieser Analyse Standardfehler verwendet, die robust gegenüber Fehlspezifikationen (Autokorrelation und Heteroskedastizität) sind. Diese robusten Schätzer der Varianz gehen zurück auf die Arbeiten von Huber (1967) und White (1980).

Zur Veranschaulichung ist dieses Datenformat beispielhaft in folgender Abbildung 7 dargestellt.

ID	Skalenlayout	Item-Werte z.B. DK-Kategorie (ja=1 / nein=0)	Item-Konstrukt (PES=0 / LIST=1)	Einstellungsstärke (gegenüber Thema des Konstrukts)
1	ADD-DK	Item 1 : 0	0	9
1	ADD-DK	Item 2: 0	0	9
1	ADD-DK	Item 3: 0	0	9
1	ADD-DK	Item 4: 0	0	9
1	ADD-DK	Item 5: 1	0	9
1	ADD-DK	Item 6: 0	0	9
1	ADD-DK	Item 7: 0	0	9
1	ADD-DK	Item 8: 1	1	9
1	ADD-DK	Item 9: 0	1	12
1	ADD-DK	Item 10: 0	1	12
1	ADD-DK	Item 11: 0	1	12
1	ADD-DK	Item 12: 0	1	12
1	ADD-DK	Item 13: 1	1	12
1	ADD-DK	Item 14: 0	1	12
1	ADD-DK	Item 15: 0	1	12
1	ADD-DK	Item 16: 1	1	12
1	ADD-DK	Item 17: 0	1	12
2	SEP-DK	Item 1 : 0	0	9
2	SEP-DK	Item 2: 0	0	9
2	SEP-DK	Item 3: 1	0	9
2	SEP-DK	Item 4: 0	0	9
2	SEP-DK	Item 5: 0	0	9
2	SEP-DK	Item 6: 1	0	9
2	SEP-DK	Item 7: 0	0	9
2	SEP-DK	Item 8: 0	1	9
2	SEP-DK	Item 9: 0	1	12
2	SEP-DK	Item 10: 1	1	12
:	:	:	:	:

Abbildung 7 Datensatz im „long-format“ (Items = Fälle)

Aufgrund dieses Vorgehens ändert sich auch die Fallzahl, von 307 Befragten auf 5219 Fälle.⁶ In der tatsächlichen Analyse liegt die Fallzahl dann etwas niedriger, da hierbei noch all jene Fälle ausgeschlossen werden, die auf mindestens einer der in

⁶ Die neue Fallzahl von 5219 ergibt sich aus der Anzahl der Befragten (N = 307) multipliziert mit der Zahl der untersuchten Items (N = 17)

dem jeweiligen Regressionsmodell berücksichtigten Variablen einen fehlenden Wert enthalten.

In der multivariaten Analyse der Studie 1 wird demnach der Datensatz im „long format“ eingesetzt (die Items werden als Fälle betrachtet), auf den schließlich die Analysemethode der logistischen Regression angewandt wird, da die abhängige Variable dichotom ist. Somit kann die mittlere Neigung über alle Befragte untersucht werden, wobei die rechte extreme, die mittlere bzw. die DK-Kategorie in Abhängigkeit davon zu wählen ist, in welchem Layout die Ratingskala präsentiert wurde, welches Konstrukt die Items repräsentieren und wie hoch die individuelle Einstellungsstärke gegenüber dem Fragethema ist.

6.3.3 Interpretation der Ergebnisse anhand marginaler Effekte

Die Interpretation der Ergebnisse der logistischen Multilevel-Regressionen mit fixen Effekten der Studie 1 erfolgt zum einen anhand sogenannter vorhergesagter Wahrscheinlichkeiten. Dieses Vorgehen wird gewählt, da Logit-Koeffizienten logistischer Regressionmodelle zwar in ihrem Vorzeichen und in ihrer Signifikanz einfach interpretierbar sind, eine weitergehende Betrachtung hinsichtlich ihrer Effektstärke und der inhaltlichen Relevanz aufgrund ihrer Einheit (logarithmiertes Wahrscheinlichkeitsverhältnis) jedoch schwierig ist (Long und Freese 2006). Zusätzlich ist der Vergleich der Koeffizienten der unabhängigen Variablen eines Modells untereinander nur dann möglich, wenn sie die gleiche Skalierung besitzen, was beispielsweise in Studie 1 für die Variablen „Skalenlayout“ und „Einstellungsstärke“ nicht der Fall ist. Daher empfehlen Long und Freese (2006) sowie Best und Wolf (2010), für eine inhaltlich bedeutsame Interpretation auf die vorhergesagten Wahrscheinlichkeiten für definierte Befragentypen oder auf Maße zurückzugreifen, die auf diesen vorhergesagten Werten basieren.⁷ Zum Beispiel ist es in Studie 1 interessant zu vergleichen, wie sich die vorhergesagten Wahrscheinlichkeiten der Wahl einer bestimmten Antwortkategorie zwischen den drei Layoutgruppen oder den beiden Item-Konstrukten unterscheiden.

⁷ Hier sei kurz erwähnt, dass auch Odds-Ratios (Chancenverhältnis) eine Alternative zu AMEs darstellen. Allerdings empfehlen Best und Wolf (2010) die Interpretation logistischer Regressionsergebnisse anhand AMEs zu favorisieren, da diese den Odds-Ratios im Hinblick auf Robustheit, Interpretierbarkeit und Additivität überlegen sind.

Zum anderen werden zur Interpretation der Analyseergebnisse der Studie 1 auch die sogenannten „marginalen Effekte“ genutzt, welche ein weiteres Maß zur Interpretation der Ergebnisse von logistischen Regressionsmodellen darstellen. Der marginale Effekt, auch als Effekt auf die Wahrscheinlichkeit bezeichnet, versucht den durchschnittlichen Einfluss der unabhängigen Variable auf die Wahrscheinlichkeit des Auftretens der untersuchten abhängigen Variable in einer einzigen Kennziffer auszudrücken. Dabei sind marginale Effekte „nicht nur abhängig vom Regressionskoeffizienten β_j , sondern zusätzlich von der Ausprägung aller Variablen und ihrem Effekt [...]. Der marginale Effekt von x_j [der unabhängigen Variablen] [variiert] erstens mit der Ausprägung von x_j selbst und zweitens mit den Ausprägungen der anderen unabhängigen Variablen.“ (Best und Wolf 2010, S.839) Um den durchschnittlichen Einfluss der unabhängigen Variablen auf die Wahrscheinlichkeit des Auftretens der abhängigen Variablen betrachten zu können, kann entweder der durchschnittliche marginale Effekt (im Englischen als „average marginal effect“ bezeichnet, kurz AME) oder der marginale Effekt am Durchschnitt (im Englischen als „marginal effect at the mean“, kurz MEM) herangezogen werden: Der Durchschnittseffekt wird bei AMEs als Mittelwert der marginalen Effekte aller Beobachtungen berechnet. Bei MEMs hingegen wird der Durchschnittseffekt am Mittelwert aller Variablen berechnet. Ein Vorteil des AME gegenüber dem MEM ist, dass der AME nicht von der unkorrelierten unbeobachteten Heterogenität verzerrt wird und daher auch für den Vergleich zwischen Modellen herangezogen werden kann (Best und Wolf 2010). Bei der Betrachtung der Regressionsergebnisse in dem Ergebnisteil der Studie 1 (vgl. Unterabschnitt 9) werden daher AMEs zur Interpretation der Zusammenhänge innerhalb und zwischen den Modellen genutzt. Neben der Robustheit gegenüber der unbeobachteten Heterogenität haben AMEs noch den weiteren Vorteil, eine intuitive Interpretation zu ermöglichen, nämlich als durchschnittlichen Effekt auf die Wahrscheinlichkeit. Demnach steigt die Wahrscheinlichkeit des Auftretens des untersuchten Zustandes um AME-Prozentpunkte an, wenn die unabhängige Variable um eine Einheit steigt. Der Nachteil des AME ist allerdings, dass es sich um einen Durchschnittseffekt handelt, der den nichtlinearen Verlauf der Wahrscheinlichkeitskurve nicht abbildet (Best und Wolf 2010).

1. Marginale Effekte nominaler Variablen

Einige abhängige Variablen dieser Untersuchung, wie beispielsweise das Skalenlayout, sind nicht metrisch, sondern besitzen ein nominales Skalenniveau. Für diese Variablen werden anstatt der AMEs die sogenannten „average discrete change effects“ ermittelt. Bei diesen Effekten wird die Veränderung der vorhergesagten Wahrscheinlichkeit bei Änderungen der interessierenden unabhängigen Variablen von einem auf einen anderen Zustand betrachtet und nicht wie bei metrischen unabhängigen Variablen, wenn deren Werte um eine Einheit ansteigen. Von den „average discrete change effects“ wird also die mittlere Differenz zwischen den vorhergesagten Wahrscheinlichkeiten der einzelnen Ausprägungen einer nominalen Variable, beispielsweise die drei Layoutgruppen, angezeigt. Wie bei den AMEs werden diese Veränderungen über alle Beobachtungen gemittelt. Diese Effekte werden üblicherweise auch als AMEs bezeichnet, obwohl sich die Berechnung unterscheidet. Demnach wird der AME hier für die drei Gruppen des Antwortskalenlayouts berechnet, indem innerhalb der einzelnen Gruppen für jede zugehörige Beobachtung der marginale Effekt unter Konstanzhaltung der übrigen unabhängigen Variablen ermittelt wird, und aus diesen marginalen Effekten ein durchschnittlicher Effekt berechnet wird (Auspurg und Hinz 2010).

2. Marginale Effekte zur Interpretation von Interaktionseffekten

Auch die Interpretation der Signifikanz und der Stärke von Interaktionseffekten ist in logistischen Modellen aus zuvor genannten Gründen anhand der Logit-Koeffizienten schwierig. Durch die Nichtlinearität kann anhand des ausgewiesenen Signifikanzniveaus des Logit-Koeffizienten eines Interaktionsterms nicht auf die Signifikanz der Interaktion für alle Ausprägungskombinationen der unabhängigen Variablen geschlossen werden. Beispielsweise wird in Studie 1 die Hypothese geprüft, ob das Layout der Ratingskala befragte Personen nur dann signifikant in der Wahl der Antwortskala beeinflusst, wenn sie eine geringe bis mittlere Einstellungsstärke gegenüber dem abgefragten Thema besitzen. Für befragte Personen mit einer hohen Einstellungsstärke hingegen wird angenommen, dass sie nicht auf Informationen des Layouts bei der Antwortgabe zurückgreifen und deshalb keine Layouteffekte für sie beobachtet werden. Diese moderierende Wirkung der Einstellungsstärke auf den Effekt des Layouts kann nicht direkt aus dem Logit-Koeffizienten des Interaktionsterms aus der Ergebnistabelle abgelesen

werden. Daher lautet auch hier die Empfehlung, die Stärke und Signifikanz von Interaktionseffekten anhand der AMEs der einzelnen Ausprägungen der unabhängigen Variablen zu interpretieren (Best und Wolf 2010).

In Studie 1 werden beispielsweise alle AMEs der möglichen Merkmalskombinationen von Antwortskalenlayout (NO-DK / SEP-DK / ADD-DK) und Einstellungsstärke (Werte von 1 bis 25 Punkten) berechnet. Da die Interpretation des Interaktionseffekts aufgrund (der in diesem Beispiel recht hohen) Anzahl der AMEs schnell unübersichtlich wird, wird in dieser Untersuchung die Interpretation anhand der graphischen Darstellung der AMEs erfolgen. Dieses Vorgehen wird hier beispielhaft anhand der in Abbildung 8 gezeigten Grafik beschrieben.

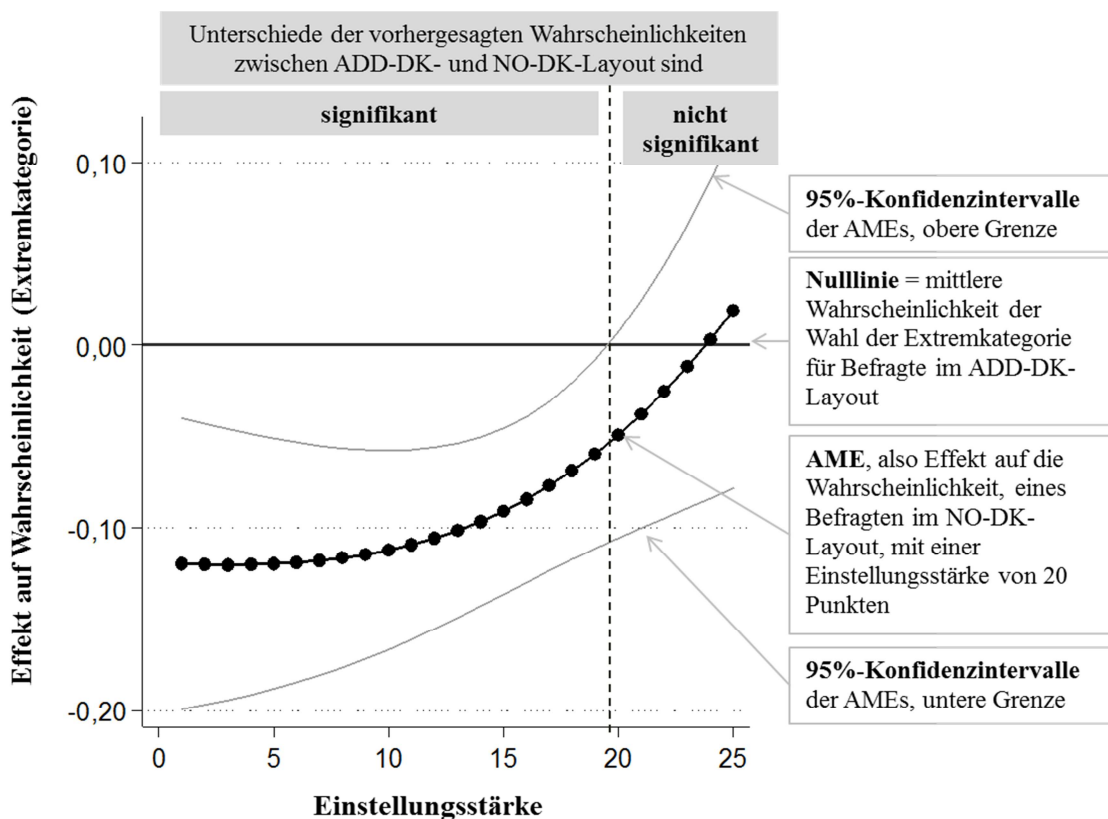


Abbildung 8 Graphik der marginalen Effekte (AMEs) von Befragten im NO-DK-Layout nach Einstellungsstärke (Referenz: ADD-DK)

Die Kurve in Abbildung 8 gibt die Effektunterschiede auf die Wahrscheinlichkeit der Wahl der Extremkatgeorie zwischen dem ADD-DK-Layout und dem NO-DK-Layout nach Einstellungsstärke wieder. Dabei wird die mittlere Wahrscheinlichkeit der Wahl der rechten Extremkategorie der Befragten, die anhand des ADD-DK-

Layouts geantwortet haben, durch die *Nulllinie* repräsentiert. Die *schwarzen Datenpunkte* der Kurve sind die AMEs der Befragten im NO-DK-Layout mit der jeweils auf der x-Achse ablesbaren Einstellungsstärke. Die *grauen Bänder* entlang der Kurve bilden die oberen und unteren Grenzen der 95 %-Konfidenzintervalle der AMEs ab. Schneiden die Konfidenzintervalle nun die Nulllinie, die ja die mittlere vorhergesagte Wahrscheinlichkeit der Befragten im ADD-DK-Layout repräsentiert, bedeutet dies, dass der Interaktionseffekt in diesem Bereich nicht signifikant ist. In Abbildung 8 ist zu sehen, dass Befragte mit einer Einstellungsstärke bis ca. 19 Punkten eine signifikant niedrigere (der AME ist negativ) Wahrscheinlichkeit besitzen als Befragte mit der jeweils selben Einstellungsstärke im ADD-DK-Layout. Ein AME von beispielsweise -0,10 Punkten kann dann als ein Unterschied von 10 Prozentpunkten, wahrscheinlich die Extremkatgorie zu wählen, zwischen den beiden Layouts interpretiert werden.

6.3.4 Aufbau der logistischen Regressionsmodelle

Für jedes der drei hier interessierenden Antwortverhalten – rechte extreme, mittlere bzw. DK-Antworten - werden jeweils fünf logistische Regressionsmodelle untersucht (vgl. Abbildung 9). Dabei beinhaltet Modell 1 lediglich eine Indikatorvariable für die experimentellen Bedingungen. Es wird also getestet, ob das Layout der Antwortskala einen Effekt auf die Wahl der Antwortkategorie ausübt, wie in den Hypothesen (H 1.1), (H 1.2) und (H 1.2) formuliert (vgl. Unterabschnitt 5.3).

In Modell 2 wird zusätzlich zu dem Skalenlayout eine weitere Indikatorvariable hinzugenommen, die jeweils als Zuweisung der siebzehn untersuchten Items zu einem der beiden betrachteten Item-Konstrukte (PES / LIST) dient. Somit wird der Einfluss des thematischen Item-Inhaltes der verschiedenen Konstrukte auf die Wahl der Antwortkategorie kontrolliert. In Modell 3 wird unter Hinzunahme eines Interaktionsterms „Skalenlayout \times Item-Konstrukt“ untersucht, ob für die beiden untersuchten Item-Konstrukte PES und LIST unterschiedliche Layouteffekte beobachtet werden können, wie in Hypothese (H 2) angenommen.

Abhängige Variablen	
1.	Wahl der Extremkategorie (Unterabschnitt 9.1)
2.	Wahl der Mittelkategorie (Unterabschnitt 9.2)
3.	Wahl der DK-Kategorie (Unterabschnitt 9.3)
Unabhängige Variablen	
Modell 1:	+ „Antwortskalenlayout“
Modell 2:	+ „Item-Konstrukt“
Modell 3:	+ Interaktion „Item-Konstrukt \times Layout“
Modell 4:	+ „Einstellungsstärke“
Modell 5:	+ Interaktion „Einstellungsstärke \times Layout“

Abbildung 9 Aufbau der geschachtelten logistischen Multilevel-Regressionen der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie

In Modell 4 wird die Einstellungsstärke in Form einer metrischen Variable als Haupteffekt in das Modell aufgenommen. Damit wird getestet, inwieweit die individuelle Einstellungsstärke Einfluss auf das Auftreten von positiven, extremen, mittleren bzw. DK-Antworten nimmt. Modell 5 erweitert schließlich das Multilevel-Modell noch um den Interaktionsterm „Layout \times Einstellungsstärke“. Somit wird die Hypothese (H 3) geprüft, die besagt, dass Effekte des Antwortskalenlayouts eher bei Befragten mit geringer als mit hoher Einstellungsstärke gegenüber dem mit der Skala erhobenen Konstrukt auftreten.

7. Deskriptive Ergebnisse

Dieser Abschnitt präsentiert die deskriptiven Ergebnisse der Studie 1. Im Folgenden werden die Antwortverteilungen der Items der PES (Unterabschnitt 7.1), des LIST-Inventars (Unterabschnitt 7.2) und der Items sowie des Indexes der Einstellungsstärke (Unterabschnitt 7.3) beschrieben. Im letzten Unterabschnitt (7.4) werden die prozentualen Häufigkeiten der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie zu den beiden Item-Konstrukten PES und LIST betrachtet.

7.1 Antwortverteilung der Political Efficacy-Skala

Die Ergebnisse der Varianzanalyse zeigen, dass keine signifikanten Unterschiede der Mittelwerte für die einzelnen PES-Items und für den summierten Index aller sieben Items zwischen den drei untersuchten Antwortskalenlayouts NO-DK-, SEP-DK- und ADD-DK-Layout bestehen (vgl. Tabelle A 9).⁸ Die Mittelwerte der PES-Items reichen im NO-DK-Layout von $M_{Item3} = 2,20$ ($SD = 0,99$) bis $M_{Item7} = 3,67$ ($SD = 0,99$). Im SEP-DK Layout werden Item-Mittelwerte in Höhe von $M_{Item5} = 2,22$ ($SD = 1,11$) bis $M_{Item7} = 3,78$ ($SD = 0,84$) beobachtet. Die arithmetischen Mittelwerte der Items im ADD-DK-Layout reichen schließlich von $M_{Item5} = 2,16$ ($SD = 1,22$) bis $M_{Item7} = 3,71$ ($SD = 1,14$). Der summierte Mittelwert über alle sieben Items beträgt im NO-DK-Layout $M = 2,89$ ($SD = 0,37$), im SEP-DK-Layout $M = 2,94$ ($SD = 0,33$) und im ADD-DK-Layout $M = 2,94$ ($SD = 0,34$) (vgl. Tabelle A 9).⁹

Die Verteilungen der sieben untersuchten PES-Items weichen bei allen drei Skalenlayouts signifikant von der Standardnormalverteilung ab. Die Werte der Schiefe der Verteilungen reichen im NO-DK-Layout von $S_{Item7} = -0,79$ ($SE = 0,23$) bis $S_{Item5} = 0,72$ ($SE = 0,24$); im SEP-DK-Layout $S_{Item7} = -0,67$ ($SE = 0,25$) bis $S_{Item3} = 0,80$ ($SE = 0,24$) und im ADD-DK-Layout $S_{Item7} = -0,65$ ($SE = 0,25$) bis $S_{Item5} = 0,90$ ($SE = 0,25$). Die Werte der Kurtosis reichen im NO-DK-Layout von $K_{Item1} = -1,31$ ($SE = 0,46$) bis $K_{Item7} = 0,46$ ($SE = 0,46$), im SEP-DK-Layout von $K_{Item1} = -1,18$ ($SE = 0,48$) bis $K_{Item3} = 0,19$ ($SE = 0,47$) und im ADD-DK-Layout von $K_{Item3} = -1,21$ ($SE = 0,49$) bis $K_{Item5} = -0,29$ ($SE = 0,49$) (vgl. Tabelle A 1 bis Tabelle A 7).

Während zwischen den einzelnen Items der PES keine signifikanten Unterschiede in den Verteilungen beobachtet werden können, zeigen sich hingegen für die Verteilung des summierten Index Unterschiede zwischen den drei Skalenlayouts. Während im ADD-DK-Layout die Verteilung einer Normalverteilung folgt (Shapiro – Wilk = 0,976, $p > 0,05$), ist dies nicht der Fall für die Verteilungen des

⁸ Alle in der Beschriftung mit Tabelle A gekennzeichneten Tabellen befinden sich in Anhang.

⁹ Der Ausschluss von Ausreißern, definiert als Fälle, die um mehr als 1,5 bis 3 Boxlängen vom 25% oder 75% Perzentil entfernt liegen, beeinflusst die Ergebnisse der Varianzanalyse nicht.

summierten Indexes im NO-DK-Layout ($\text{Shapiro} - \text{Wilk}_{\text{NODK}} = 0,971, p < 0,01$) und SEP-DK-Layout ($\text{Shapiro} - \text{Wilk}_{\text{SEP DK}} = 0,971, p < 0,05$) (vgl. Tabelle A 8).

Die deskriptiven Ergebnisse zeigen, dass zwischen den Verteilungen der einzelnen Items der PES sowie der Verteilungen des summierten Item-Indexes keine signifikanten Unterschiede zwischen dem NO-DK-, dem SEP-DK- und dem ADD-DK-Layout bestehen. Einzig in Bezug auf die Normalverteilung des summierten Item-Indexes wird für das ADD-DK-Layout eine bessere Anpassung im Vergleich zu den beiden anderen Layouts beobachtet.

7.2 Antwortverteilung der Items zur Erfassung von Lernstrategien

Eine Varianzanalyse der zehn LIST-Items sowie des summierten Indexes zeigt, dass weder für die einzelnen Items noch für den Index signifikante Unterschiede der Mittelwerte zwischen den drei untersuchten Antwortskalenlayouts NO-DK-, SEP-DK- und ADD-DK-Layout bestehen (vgl. Tabelle A 21). Die arithmetischen Mittelwerte der LIST-Items reichen im NO-DK-Layout von $M_{\text{Item } 3} = 2,19$ ($SD = 1,06$) bis $M_{\text{Item } 5} = 3,12$ ($SD = 0,88$). Im SEP-DK-Layout werden Item-Mittelwerte in Höhe von $M_{\text{Item } 3} = 2,26$ ($SD = 1,03$) bis $M_{\text{Item } 5} = 3,40$ ($SD = 1,04$) beobachtet. Die Mittelwerte der Items im ADD-DK-Layout reichen schließlich von $M_{\text{Item } 3} = 2,16$ ($SD = 1,14$) bis $M_{\text{Item } 5} = 3,32$ ($SD = 1,07$). Der summierte Mittelwert über alle sieben Items beträgt im NO-DK-Layout $M = 2,58$ ($SD = 0,54$), im SEP-DK-Layout $M = 2,74$ ($SD = 0,62$) und im ADD-DK-Layout $M = 2,68$ ($SD = 0,57$) (vgl. Tabelle A 21).¹⁰

Die Verteilungen der zehn untersuchten LIST-Items weichen in allen drei Skalenlayouts signifikant von der Standardnormalverteilung ab. Die Werte der Schiefe der Verteilungen reichen im NO-DK-Layout von $S_{\text{Item } 2} = -0,90$ ($SE = 0,23$) bis $S_{\text{Item } 3} = 0,84$ ($SE = 0,23$); im SEP-DK-Layout $S_{\text{Item } 2} = -0,12$ ($SE = 0,24$) bis $S_{\text{Item } 8} = 0,46$ ($SE = 0,24$) und im ADD-DK-Layout $S_{\text{Item } 2} = -0,14$ ($SE = 0,25$) bis $S_{\text{Item } 8} = 0,93$ ($SE = 0,25$). Die Werte der Kurtosis

¹⁰ Der Ausschluss von Ausreißern, definiert als Fälle, die um mehr als 1,5 bis 3 Boxlängen vom 25% oder 75% Perzentil entfernt liegen, beeinflusst die Ergebnisse der Varianzanalyse nicht.

reichen im NO-DK-Layout von $K_{Item\ 2} = -0,90$ ($SE = 0,46$) bis $K_{Item\ 8} = 0,77$ ($SE = 0,47$), im SEP-DK-Layout von $K_{Item\ 2} = -1,18$ ($SE = 0,47$) bis $K_{Item\ 2} = -0,08$ ($SE = 0,47$) und im ADD-DK-Layout von $K_{Item\ 2} = -1,12$ ($SE = 0,49$) bis $K_{Item\ 8} = 1,02$ ($SE = 0,49$) (vgl. Tabelle A 10 bis Tabelle A 19).

Anders als die einzelnen LIST-Items folgt der summierte Item-Index für alle drei Skalenlayouts der Normalverteilung (Shapiro – Wilk_{NODK} = 0,986, $p > 0,05$, Shapiro – Wilk_{SEPDK} = 0,984, $p > 0,05$, Shapiro – Wilk_{ADDK} = 0,987, $p > 0,05$) (vgl. Tabelle A 20).

Die deskriptiven Ergebnisse zeigen, dass zwischen den Verteilungen der einzelnen Items des LIST-Inventars sowie der Verteilungen des summierten Item-Indexes keine signifikanten Unterschiede zwischen dem NO-DK-, dem SEP-DK- und dem ADD-DK-Layout bestehen.

7.3 Antwortverteilungen der Items und des Indexes der Einstellungsstärke

Die individuelle Einstellungsstärke wurde mittels der beiden Items Sicherheit (bei der Fragebeantwortung) und Wichtigkeit (des Themas) jeweils direkt im Anschluss an die Items der beiden Konstrukte PES und LIST erhoben (vgl. Unterabschnitt 6.2). Die Antwortverteilung des Items Sicherheit (vgl. Tabelle 1) zeigt wie in Hypothese (H 2) angenommen (vgl. Unterabschnitt 5.3), dass sich mehr Befragte bei Beantwortung der Fragen zu den eigenen Lernstrategien (LIST-Items) als zu Fragen über die Einschätzung der politischen Selbstwirksamkeit (PES-Items) „ziemlich sicher“ (61,4 % vs. 52,2 %) oder „sehr sicher“ (25,2 % vs. 14,6 %) gefühlt haben. Dies spiegelt sich auch in einem signifikant höheren Mittelwert für die Sicherheit bei Beantwortung der LIST-Items wider ($t = 6,28$; $df = 300$). Hingegen „wenig sicher“ oder gar „überhaupt nicht sicher“ haben sich bei Beantwortung der Items beider Konstrukte nur sehr wenige Befragte gefühlt, weshalb die beiden Mittelwerte mit $M_{PES} = 3,75$ und $M_{LIST} = 4,10$ eher hoch sind (vgl. Tabelle 1).

Tabelle 1 Antwortverteilungen des Items Sicherheit bei der Beantwortung der Items zur politischen Selbstwirksamkeit (PES) und zu den eigenen Lernstrategien (LIST)

	PES	LIST
(1) überhaupt nicht sicher	1,0 %	0,7 %
(2) weniger sicher	4,7 %	1,6 %
(3) mittelmäßig sicher	27,6 %	11,1 %
(4) ziemlich sicher	52,2 %	61,4 %
(5) sehr sicher	14,6 %	25,2 %
N	301	306
Mittelwert (SD)	3,75 (0,80)	4,10 (0,69)

Dem Thema der politischen Selbstwirksamkeit sprechen die Befragten eine leicht höhere persönliche Wichtigkeit zu als den persönlichen Lernstrategien (vgl. Tabelle 2). Die beiden Mittelwerte unterscheiden sich signifikant voneinander ($t = 3,23$; $df = 303$), zeigen aber mit Werten $M_{PES} = 3,96$ und $M_{LIST} = 3,76$, dass beide Themen gleichermaßen eine eher hohe Wichtigkeit für die Befragten besitzen.

Tabelle 2 Antwortverteilungen des Items „Wichtigkeit“ des Themas der politischen Selbstwirksamkeit (PES) und der eigenen Lernstrategien (LIST)

	Politische Selbstwirksamkeit (PES-Items)	Eigene Lernstrategien (LIST-Items)
(1) überhaupt nicht wichtig	0,7 %	1,6 %
(2) weniger wichtig	3,3 %	7,2 %
(3) mittelmäßig wichtig	22,5 %	23,7 %
(4) ziemlich wichtig	46,7 %	48,7 %
(5) sehr wichtig	26,8 %	18,8 %
N	306	304
Mittelwert (SD)	3,96 (,047)	3,76 (,051)

Die Randomisierung hinsichtlich der Items der Einstellungsstärke war erfolgreich; Chi-Quadrat-Tests weisen als Ergebnis aus, dass sich die Verteilungen der Items Sicherheit und Wichtigkeit nicht signifikant zwischen den drei Layoutgruppen unterscheiden (vgl. Tabelle A 22 bis Tabelle A 25). Auch für die Item-Mittelwerte werden in einer Varianzanalyse keine signifikanten Unterschiede zwischen den drei Layouts gefunden (vgl. Tabelle A 26).

Somit wird die Einstellungsstärke eines Befragten gegenüber den beiden Konstrukten PES und LIST jeweils als multiplikativer Index gebildet (vgl. auch Unterabschnitt 6.2). Für die Einstellungsstärke gegenüber der politischen

Selbstwirksamkeit reichen die Werte von 4 bis 25 Punkte, der Mittelwert beträgt $M = 14,99$ ($SD = 4,943$). Für die Einstellungsstärke gegenüber den Lernstrategien beträgt der Minimalwert 3 und der Maximalwert 25, der Mittelwert liegt bei $M = 15,48$ ($SD = 5,07$). Die beiden Mittelwerte der Einstellungsstärken gegenüber den beiden Item-Konstrukten unterscheiden sich nicht signifikant voneinander ($t = 1,36$; $df = 298$).

7.4 Extreme, mittlere und Don't Know-Antworten in den Item-Konstrukten

Bei Beantwortung der sieben Items der PES wurde die rechte Extremkategorie „stimme stark zu“ von 48,9 % der Befragten zur Antwortgabe gewählt; 51,1 % der Befragten machten keinen Gebrauch von diese Kategorie zur Beantwortung der PES-Items. Die Mittelkategorie „weder noch“ wurde von 69,1 % der Befragten gewählt, die übrigen 30,9 % nutzten diese Kategorie nicht für die PES-Items. Die DK-Kategorie schließlich wurde bei den PES-Items von 18 % der Befragten zur Antwortgabe genutzt und von 82 % der Befragten nicht genutzt (vgl. Tabelle 3).

Tabelle 3 Auswahlhäufigkeiten der rechten Extrem-, Mittel- und DK-Kategorie in der PES

	rechte Extrem- kategorie	Mittel- kategorie	DK- Kategorie
Gewählt „nein“	51,1 %	29,6 %	82,0 %
Gewählt „ja“	48,9 %	70,4 %	18,0 %
	100 %	100 %	100 %
	(N = 307)	(N = 307)	(N = 200)
<i>wenn gewählt (=ja), dann zur Antwortabgabe von</i>			
...1 Item	73,3 %	57,4 %	75,0 %
...2 Items	22,0 %	25,9 %	16,7 %
...3 Items	4,7 %	13,0	8,3 %
...4 Items	-	0,5 %	-
...5 Items	-	-	-
...6 Items	-	-	-
...7 Items	-	-	-
	100 %	100 %	100 %
	(N = 150)	(N = 216)	(N = 36)

Eine Betrachtung der Auswahlhäufigkeiten der Befragten, welche die positive Extremkategorie am rechten Skalenende zur Antwortgabe nutzten, zeigt, dass die

meisten – nämlich 73,3 % der Befragten – diese Kategorie nur für eines der sieben PES-Items wählten, weitere 22,0 % der Befragten beantworteten zwei Items und 4,7 % der Befragten drei Items mit der rechten Extremkategorie. Auch die DK-Kategorie wurde von den meisten Befragten, die Gebrauch von dieser Kategorie zur Beantwortung der PES-Items machten, nur für ein Item genutzt (75,0 %), einige (16,7 %) beantworteten anhand dieser Kategorie zwei Items und wenige (8,3 %) drei Items der PES. Auf Basis dieser Auswahlhäufigkeiten der rechten Extrem- und der DK-Kategorie wird nicht von einer Antworttendenz anhand dieser beider Kategorien ausgegangen. Bei der Betrachtung der Auswahlhäufigkeiten der Mittelkategorie ist zu sehen, dass ein Teil der Befragten, die diese Kategorie zur Antwortgabe nutzten, diese für mehr als die Hälfte der PES-Items – nämlich viermal – wählten. Allerdings sind dies gerade mal 0,5 % der Befragten (vgl. Tabelle 3).

Bei Beantwortung der zehn Items des LIST-Inventars wurde die rechte Extremkategorie „trifft voll und ganz zu“ von 69,1 % der Befragten nicht gewählt und von 30,9 % der Befragten gewählt. Die Mittelkategorie „trifft teilweise zu“ wurde von 88,3 % der Befragten genutzt und von 11,7 % der Befragten nicht genutzt. Die DK-Kategorie wählten schließlich 95,0 % der Befragten nicht zur Beantwortung der zehn LIST-Items aus, 5,0 % der Befragten nutzten diese Kategorie (vgl. Tabelle 4).

Die Auswahlhäufigkeiten zeigen für die rechte Extrem- und die DK-Kategorie wiederum kein Antwortverhalten, das auf eine systematische Antworttendenz hinweisen würde. Die überwiegende Mehrheit der Befragten, welche die rechte Extremkategorie (55,8 %) oder die DK-Kategorie (60 %) auswählten, tat dies lediglich für ein Item. Anders ist dies bei der Mittelkategorie. Ein nicht geringer Anteil von 14,0 % der Befragten, die anhand der Mittelkategorie die LIST-Items beantworteten, nutzte diese Kategorie für sechs bis hin zu allen zehn LIST-Items (vgl. Tabelle 4). Da jedoch bei der unipolaren Antwortskala der LIST-Items die Mittelkategorie eine teilweise zutreffende Beschreibung der persönlichen Lernsituation ausdrückt, also eine inhaltliche Bedeutung hat, und die Items ähnliche Aspekte der gleichen Dimension beschreiben, zeigt eine so häufige Nutzung der Mittelkategorie nicht zwingend eine Einstellung, die nicht auch empirisch

vorstellbar ist. Daher werden keine Befragten aus der späteren Analyse ausgeschlossen.

Tabelle 4 Auswahlhäufigkeiten der rechten Extrem-, Mittel- und DK-Kategorie für LIST

	rechte Extrem- kategorie	Mittel- kategorie	DK- Kategorie
Gewählt „nein“	69,1 %	11,7 %	95,0 %
Gewählt „ja“	30,9 %	88,3 %	5,0 %
	100 % (N = 307)	100 % (N = 307)	100 % (N = 200)
<i>wenn gewählt (=ja), dann zur Antwortabgabe von</i>			
...1 Item	55,8 %	22,1 %	60,0 %
...2 Items	18,9 %	20,7 %	20,0 %
...3 Items	9,5 %	17,0 %	-
...4 Items	7,4 %	18,8 %	10,0 %
...5 Items	4,2 %	9,2 %	-
...6 Items	4,2 %	8,5 %	10,0 %
...7 Items	-	2,2 %	-
...8 Items	-	0,7 %	-
...9 Items	-	0,4 %	-
...10 Items	-	0,4 %	-
	100 % (N = 95)	100 % (N = 271)	100 % (N = 10)

8. Bivariate Ergebnisse

Dieser Abschnitt behandelt eine erste, bivariate Analyse der Zusammenhänge zwischen der Wahl der rechten extremen, der mittleren und der DK-Kategorie und dem Ratingskalenlayout. Dabei wird zuerst die Wahl der drei hier untersuchten Antwortkategorien ausschließlich in Abhängigkeit von dem Skalenlayout betrachtet (Unterabschnitt 8.1). Anschließend wird zusätzlich noch die Einstellungsstärke – als dichotome Variable mit den Ausprägungen niedrig / mittel versus hoch - als Einflussgröße auf die Wahl der Antwortkategorie in die Analyse miteinbezogen (Unterabschnitt 8.2). Wie in der Vorstellung der Studie 1 in Unterabschnitt (5.1) bereits beschrieben, konzentriert sich diese Untersuchung nicht auf Effekte des Antwortskalenlayouts auf Item- sondern auf Skalenebene. Daher wird die Wahl der

rechten Extrem-, der Mittel- und der DK-Kategorie über jeweils alle Items der beiden Item-Konstrukte PES und LIST hinweg betrachtet. Es wird demnach von der Wahl der jeweiligen Antwortkategorie gesprochen, wenn ein Befragter für mindestens eines der sieben PES-Items beziehungsweise eines der zehn LIST-Items die untersuchte Kategorie zur Beantwortung gewählt hat.

8.1 Kategoriewahl in Abhängigkeit von Skalenlayout

Rechte Extremkategorie

Zwischen den drei Layoutgruppen sind deutliche Unterschiede hinsichtlich der Wahl der rechten Extremkategorie über die sieben PES-Items festzustellen (Tabelle 5): Im ADD-DK-Layout wählten mit 64,9 % mit Abstand die meisten Befragten die rechte Extremkategorie zur Beantwortung mindestens eines der PES-Items. Im NO-DK-Layout sind es 26,6 Prozentpunkte und im SEP-DK-Layout 20,2 Prozentpunkte weniger Befragte. Nach dem Chi-Quadrat-Test ist dieser Unterschied signifikant.

Bei den LIST-Items sind ebenfalls Unterschiede in gleicher Richtung, aber in geringerem Ausmaß hinsichtlich der Wahl der rechten Extremkategorie zu beobachten (vgl. Tabelle 5). Wieder wählten im ADD-DK-Layout die meisten Befragten, nämlich 38,9 %, die rechte Extremkategorie für mindestens eines der zehn LIST-Items aus; im NO-DK-Layout sind es 15,7 Prozentpunkte und im SEP-DK-Layout 5,1 Prozentpunkte weniger. Auch dieser Unterschied ist nach dem Chi-Quadrat-Test signifikant.

Diese Ergebnisse unterstützen die Hypothese (H 1.1), dass zum einen die Verschiebung des visuellen Mittelpunktes und zum anderen das nicht deutlich erkennbare Skalenende am rechten Rand im ADD-DK-Layout weitaus mehr Befragte die rechte Extremkategorie wählen lässt, als es in den beiden anderen Layouts der Fall ist. Der Unterschied zwischen dem NO-DK- und SEP-DK-Layout legt den Schluss nahe, dass durch die angebotene, wenn auch abgetrennte DK-Kategorie, die Skala im SEP-DK-Layout breiter erscheint als im NO-DK-Layout und dadurch die Extremkategorie am rechten Skalenrand in der Wahrnehmung der Befragten etwas an Extremität verliert. Dies kann dazu führen, dass mehr Befragte

diese Kategorie im SEP-DK- wählen als im NO-DK-Layout (vgl. Unterabschnitt 5.3).

Tabelle 5 Wahl der rechten Extrem-, Mittel- und DK-Kategorie in den Item-Konstrukten nach Skalenlayout

	NO-DK	Layout SEP-DK	ADD-DK
PES-Items*			
rechte Extremkategorie	38,3 %	44,7 %	64,9 %
Mittelkategorie	77,6 %	70,9 %	61,9 %
DK-Kategorie	-	23,3 %	12,4 %
LIST-Items*			
rechte Extremkategorie	22,4 %	33,0 %	38,9 %
Mittelkategorie	91,6 %	92,2 %	80,4 %
DK-Kategorie	-	5,8 %	4,1 %
N	107	103	97
PES		LIST	
$\chi^2_{Ex. Kat.} = 15,53, df 2, p = .000$		$\chi^2_{Ex. Kat.} = 6,189, df 2, p = .045$	
$\chi^2_{Mit. Kat.} = 6,044, df 2, p = .049$		$\chi^2_{Mit. Kat.} = 8,487, df 2, p = .014$	
$\chi^2_{DK Kat.} = 4,043, df 1, p = .044$		$\chi^2_{DK Kat.} = 0,0304, df 1, p = .581$	

* Die hier aufgeführten Prozentwerte geben an, wie viele Befragte einer Layoutgruppe die jeweilige Antwortkategorie für mindestens eines der sieben Items der PES bzw. eines der zehn Items des LIST-Inventars zur Antwortabgabe gewählt haben.

Auch die Hypothese (H 2) hinsichtlich des Auftretens von Layouteffekten in den beiden Item-Konstrukten wird durch die Ergebnisse unterstützt. Die geringeren Unterschiede zwischen den drei Layoutgruppen in der Wahl der rechten Extremkategorie bei Beantwortung der LIST-Items lässt darauf schließen, dass Befragte bei Beantwortung der Items über ihre Lernstrategien weniger empfänglich für das Skalenlayout sind, da sie die Antwort aus dem Gedächtnis abrufen können und nicht erst in einem aufwendigeren kognitiven Prozess, wie dies für die PES-Items erwartet wird, generieren müssen (vgl. Unterabschnitt 5.3).

Mittelkategorie

Für die Beantwortung der PES-Items zeigt der Vergleich über die drei Skalenlayouts, dass die Mittelkategorie am häufigsten im NO-DK-Layout gewählt wird (vgl. Tabelle 5). Hier liegt die Rate bei 77,6 % der Befragten. Im ADD-DK-

Layout hingegen nutzen gut 15 %-Punkte weniger, nämlich 61,9 % der Befragten diese Kategorie zur Antwortgabe. Befragte im SEP-DK-Layout liegen mit einer Rate von 70,9 % dazwischen. Diese Verteilungen der Wahl der Mittelkategorie nach Antwortskalenlayout unterscheiden sich gemäß dem Chi-Quadrat-Test signifikant voneinander.

Bei Beantwortung der LIST-Items ist zu sehen, dass die Mittelkategorie im NO-DK-Layout mit 91,6 % und im SEP-DK-Layout mit 92,2 % von nahezu gleich vielen Befragten gewählt wird (vgl. Tabelle 5). Im ADD-DK-Layout ist diese Rate mit 80,4 % der Befragten niedriger, aber immer noch auf einem hohen Niveau. Auch für die LIST-Items weist der Chi-Quadrat-Test einen signifikanten Unterschied aus.

Das Ergebnis für die LIST-Items lässt darauf schließen, dass, wie in Hypothese (H 1.2) angenommen, Befragte in der Wahl der Mittelkategorie durch das Skalenlayout systematisch beeinflusst werden. Die Übereinstimmung von inhaltlichem und visuellem Mittelpunkt im NO-DK- und SEP-DK-Layout lässt Befragte dort die Mittelkategorie als solche klar erkennen und damit häufiger wählen; sei es, weil diese Kategorie ihrer wahren Einstellung entspricht oder aus anderen Gründen, wie beispielsweise Satisficing (vgl. Unterabschnitt 5.3). Interessant ist hingegen, dass bei den PES-Items, nicht wie erwartet, lediglich ein Unterschied zwischen dem NO-DK- und ADD-DK-Layout, sondern auch zwischen dem NO-DK- und SEP-DK-Layout beobachtet wird. Eine Erklärung dieses Befunds kann der Unterschied in der Vorgabe bzw. Nicht-Vorgabe der DK-Kategorie dieser beider Layouts sein. Sturgis, Roberts und Smith (2014) zeigen in ihrer Studie, dass in Ratingskalen ohne DK-Kategorie Befragte mit einer fehlenden Einstellung eher auf die Mittelkategorie ausweichen als die fehlende Einstellung frei zu berichten. Übertragen auf die Ergebnisse der Studie 1 dieser Dissertation wird daher angenommen, dass Befragte ohne Einstellung gegenüber den Inhalten der PES-Items im NO-DK-Layout eher die Mittelkategorie wählen, als die Frage unbeantwortet zu lassen. Im SEP-DK-Layout hingegen bekommen die Befragten eine DK-Kategorie klar erkennbar vorgegeben, weshalb Befragte ohne Einstellung zu dem Thema anscheinend eher die DK-Kategorie wählen, als auf die Mittelkategorie auszuweichen. Wie bei der Betrachtung der Wahl der DK-Kategorie

im Folgenden noch beschrieben wird, liegt generell die Rate an DK-Antworten bei den PES-Items deutlich höher als bei den LIST-Items. Dies weist darauf hin, dass Befragte gegenüber den Inhalten der PES-Items weniger klare Einstellungen besitzen oder dass sie ein stärkeres Satisficing-Verhalten zeigen. Daher kann es für Befragte bei Beantwortung der PES-Items im Vergleich zu den LIST-Items eher dazu kommen, dass sie im NO-DK-Layout die Mittelkategorie als Ersatz für die fehlende DK-Kategorie nutzen, entweder aus Gründen unklarer/fehlender Einstellungen oder wegen eines Satisficing Verhaltens.

In beiden Item-Konstrukten werden Layouteffekte beobachtet. Wieder sind diese bei den PES-Items etwas stärker ausgeprägt, was die Annahme der Hypothese (H 2) unterstützt (vgl. Unterabschnitt 5.3).

Schließlich kann als Erklärung für die insgesamt höheren Raten an mittleren Antworten zu den LIST- als zu den PES-Items in den drei Skalenlayouts die unterschiedlichen Dimensionen der beiden Antwortskalen angeführt werden. Die Antwortskala der LIST-Items ist unipolar. Der Befragte soll darauf abstufen, inwieweit die verschiedenen Lernsituationen auf ihn „überhaupt nicht“ bis „voll und ganz“ zutreffen. Die Mittelkategorie steht für ein teilweises Zutreffen der beschriebenen Lernsituation. Bei den PES-Items hingegen wird den Befragten eine bipolare Antwortskala vorgelegt. Auf dieser sollen sie eine Einstufung vornehmen, inwieweit sie die verschiedenen Aussagen über die politische Einflussnahme stark ablehnen bis ihnen stark zustimmen. Auf dieser Antwortskala steht die Mittelkategorie für eine indifferente Position („weder noch“). Es ist anzunehmen, dass die meisten Befragten für die einzelnen Items der PES eine Einstellung besitzen, die entweder eher in Richtung Ablehnung oder Zustimmung neigt und sie daher keine indifferente Position über die Mittelkategorie zum Ausdruck bringen.

DK-Kategorie

Für die letzte hier untersuchte, die DK-Kategorie, zeigt sich bei den PES-Items, dass im SEP-DK-Layout die abgetrennte DK-Kategorie von 23,3 % der Befragten für mindestens eines der sieben PES-Items zur Antwortgabe gewählt wird. Im

ADD-DK-Layout ist diese Rate mit 12,4 % der Befragten lediglich knapp halb so hoch (vgl. Tabelle 5), der Unterschied ist nach dem Chi-Quadrat-Test signifikant.

Bei den LIST-Items hingegen kann kein signifikanter Unterschied in der Wahl der DK-Kategorie zwischen den beiden Layouts beobachtet werden. Im SEP-DK-Layout geben 5,8 % und im ADD-DK-Layout 4,1 % der Befragten mindestens einmal für die zehn Konstrukt-Items eine DK-Antwort (vgl. Tabelle 5).

Aufgrund dieser Ergebnisse wird die Schlussfolgerung gezogen, dass Befragte bei Beantwortung der PES-Items stark durch das Layout der Antwortskala beeinflusst werden: Wird eine DK-Kategorie visuell mittels eines Trennstrichs von der Ratingskala abgetrennt, wählen, wie in Hypothese (H 1.3) angenommen, deutlich mehr Befragte diese Kategorie zur Antwortgabe der PES-Items als in einem Layout, bei dem die DK-Kategorie unauffällig an die Ratingskala angehängt wurde. Bei Beantwortung der LIST-Items hingegen sind die Befragten bei der Wahl der DK-Kategorie vom Skalenlayout nicht beeinflusst. Dieser Unterschied zwischen den Effekten des Skalenlayouts bei den PES- und LIST-Items unterstützt die Hypothese (H 2), die besagt, dass Layouteffekte stärker bei den PES-Items auftreten, die eher abstrakte Inhalte abfragen, zu denen Befragte ihre Einstellung noch in einem kognitiv aufwändigen Prozess konstruieren müssen. Während dieser Konstruktion greifen sie zurück auf den Kontext der Frage, hier das Skalenlayout. Zu Items hingegen, die wie im LIST-Inventar Situationen aus dem Lebensalltag behandeln, haben Befragte schnell abrufbare Informationen im Gedächtnis vorliegen, die eine Beantwortung der Items ohne hohen kognitiven Aufwand, und somit ohne die Notwendigkeit von der Einbeziehung weiterer (Kontext-) Informationen, ermöglichen (vgl. Unterabschnitt 5.3).

8.2 Kategoriewahl in Abhängigkeit von Skalenlayout und Einstellungsstärke

Eines der zentralen Studienziele ist, den Einfluss der Einstellungsstärke auf das Auftreten der Effekte des Skalenlayouts zu untersuchen. Die Annahme aus Hypothese (H 3) lautet, dass Befragte mit einer hohen Einstellungsstärke nicht so stark vom Kontext der Frage beeinflusst werden und als Folge dessen Effekte des Skalenlayouts deutlich weniger auftreten als bei Befragten mit einer niedrigen

Einstellungsstärke (vgl. Unterabschnitt 5.3). Daher werden in diesem Abschnitt die Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie in den beiden Item-Konstrukten getrennt für Befragte mit einer niedrigen bis mittleren Einstellungsstärke und für Befragte mit einer hohen Einstellungsstärke betrachtet. Hier wird nicht die metrische Variable sondern eine daraus abgeleitete, dichotome Variable der Einstellungsstärke in der Analyse eingesetzt. Diese dichotome Variable unterscheidet zwischen Befragten mit einer niedrigen bis mittleren Einstellungsstärke auf der einen und solchen mit einer hohen Einstellungsstärke auf der anderen Seite. Diese Dichotomisierung ist notwendig, weil bei dem Verfahren der Kreuztabellenanalyse eine Mindestbesetzung der Zellen vorausgesetzt wird. Diese Voraussetzung ist bei der metrischen Einstellungsstärke-Variable aber nicht erfüllt, die Fallzahlen im unteren Bereich der Einstellungsstärke sind sehr niedrig oder gleich Null, besonders wenn die Betrachtung noch nach Skalenlayout getrennt erfolgt.

Von einer hohen Einstellungsstärke gegenüber den Inhalten der PES beziehungsweise des LIST-Inventars wird dann gesprochen, wenn Befragte bei Beantwortung der beiden Items zur Erfassung der Einstellungsstärke „Sicherheit“ und „Wichtigkeit“ die vierte oder fünfte Antwortkategorie gewählt haben; sich also ziemlich bis sehr sicher gefühlt haben und das Thema ziemlich oder sehr wichtig für sie persönlich ist. Damit ergibt sich für Befragte mit einer hohen Einstellungsstärke auf dem multiplikativen Index ein Wert von 16 bis 25 Punkten. Wird auf dem Einstellungsstärke-Index ein geringerer Wert als 16 Punkte erreicht, weist dies darauf hin, dass Befragte sich entweder bei der Antwortgabe nicht sicher waren, oder dass sie dem Thema eine eher geringe persönliche Wichtigkeit zuschreiben. Daher wird bei Indexwerten von 1 bis 15 Punkten von einer niedrigen bis mittleren Einstellungsstärke gesprochen.

Die Betrachtung der Einstellungsstärke gegenüber den Inhalten des PES-Konstrukts zeigt, dass in allen drei Layoutgruppen etwas mehr als die Hälfte der Befragten eine hohe Einstellungsstärke besitzt. Gegenüber den Inhalten des LIST-Konstrukts bringt ein geringfügig größerer Anteil, d.h. in jeder Layoutgruppe etwa zwei Drittel der Befragten, eine hohe Einstellungsstärke zum Ausdruck (vgl. Tabelle A 27).

Rechte Extremkategorie

Die rechte Extremkategorie wird bei den PES-Items von sehr unterschiedlichen Anteilen der Befragten gewählt, nämlich je nachdem welches Ratingskalenlayout ihnen vorliegt und ob sie eine niedrige / mittlere oder hohe Einstellungsstärke gegenüber dem Thema der politischen Selbstwirksamkeit besitzen. Insgesamt betrachtet wählen in jeder Layoutgruppe anteilig mehr Befragte mit einer hohen Einstellungsstärke die rechte Extremkategorie zur Antwortgabe als Befragte mit einer niedrigen / mittleren Einstellungsstärke (vgl. Tabelle 6). Das ist inhaltlich sinnvoll, denn je stärker die Einstellungsstärke ist, desto eher sollten Befragte Antwortkategorien am Skalenrand zur Beantwortung der Items heranziehen, die ihre Ablehnung oder Zustimmung zu den Aussagen der Items ausdrücken.

Innerhalb der Gruppe der Befragten mit einer niedrigen / mittleren Einstellungsstärke gegenüber der politischen Selbstwirksamkeit sind ausgeprägte Unterschiede zwischen den drei Skalenlayouts zu beobachten (vgl. Tabelle 6): Im ADD-DK-Layout wählen 56,5 % der Befragten mit einer niedrigen / mittleren Einstellungsstärke die rechte Extremkategorie. Im NO-DK-Layout sind es lediglich 14,9 % der Befragten. Das ist ein Unterschied von über 40 %-Punkten. Im SEP-DK-Layout liegt die Rate mit 30,6 % der Befragten zwischen den Raten der anderen beiden Layouts und unterscheidet sich somit ebenfalls deutlich von diesen beiden. Diese Unterschiede sind nach dem Chi-Quadrat-Test signifikant.

In der Gruppe der Befragten mit einer hohen Einstellungsstärke wird ebenfalls im ADD-DK-Layout anteilig von den meisten Befragten (72,9 %) die rechte Extremkategorie zur Beantwortung mindestens eines der PES-Items gewählt. Die Anteile im NO-DK-Layout (57,6 %) und im SEP-DK-Layout (59,6 %) sind deutlich geringer und unterscheiden sich nicht untereinander (vgl. Tabelle 6). Allerdings sind die Unterschiede in den Anteilen der Befragten, die die rechte Extremkategorie wählen nicht mehr so groß wie in der Gruppe der Befragten mit einer niedrigen / mittleren Einstellungsstärke und nach dem Chi-Quadrat-Test auch nicht signifikant unterschiedlich zueinander.

Bei den LIST-Items zeigt sich, anders als bei den PES-Items, weder zwischen den Befragten mit einer niedrigen / mittleren und einer hohen Einstellungsstärke noch zwischen den drei Layoutgruppen ein klar erkennbarer Zusammenhang. Lediglich

im NO-DK-Layout wählen wie erwartet mehr Befragte mit einer hohen Einstellungsstärke (25,0 %) die rechte Extremkategorie als solche mit einer niedrigen / mittleren Einstellungsstärke (18,6 %) (vgl. Tabelle 6).

Tabelle 6 Wahl der rechten Extrem-, Mittel- und DK-Kategorie in den Item-Konstrukten nach Skalenlayout und Einstellungsstärke

	NO-DK		Layout SEP-DK		ADD-DK	
	<i>ES⁺ niedrig- mittel</i>	<i>ES⁺ hoch</i>	<i>ES⁺ niedrig- mittel</i>	<i>ES⁺ hoch</i>	<i>ES⁺ niedrig- mittel</i>	<i>ES⁺ hoch</i>
PES-Items*						
rechte Extremkategorie	14,9 %	57,6 %	30,6 %	59,6 %	56,5 %	72,9 %
Mittelkategorie	85,1 %	71,2 %	75,5 %	64,4 %	67,4 %	56,3 %
DK-Kategorie	-	-	30,6 %	17,3 %	8,7 %	16,7 %
N	47	59	49	52	46	48
LIST-Items*						
rechte Extremkategorie	18,6 %	25,0 %	38,1 %	30,0 %	37,8 %	39,7 %
Mittelkategorie	93,0 %	90,6 %	92,9 %	91,7 %	81,1 %	82,8 %
DK-Kategorie	-	-	2,4 %	8,3 %	5,4 %	3,4 %
N	43	64	42	60	37	58

PES-Items

Niedrige / mittlere Einstellungsstärke:

$$\chi^2_{Ex.Kat.} = 18,344, df 2, p = .000$$

$$\chi^2_{Mit.Kat.} = 4,018, df 2, p = .134$$

$$\chi^2_{DK-Kat.} = 7,123, df 1, p = .008$$

Hohe Einstellungsstärke:

$$\chi^2_{Ex.Kat.} = 3,007, df 2, p = .222$$

$$\chi^2_{Mit.Kat.} = 2,6, df 2, p = .272$$

$$\chi^2_{DK-Kat.} = 0,007, df 1, p = .932$$

LIST-Items

Niedrige / mittlere Einstellungsstärke:

$$\chi^2_{Ex.Kat.} = 4,872, df 2, p = .088$$

$$\chi^2_{Mit.Kat.} = 3,809, df 2, p = .149$$

$$\chi^2_{DK-Kat.} = 0,493, df 1, p = .483$$

Hohe Einstellungsstärke:

$$\chi^2_{Ex.Kat.} = 3,110, df 2, p = .211$$

$$\chi^2_{Mit.Kat.} = 2,745, df 2, p = .253$$

$$\chi^2_{DK-Kat.} = 1,261, df 1, p = .261$$

+ ES= Einstellungsstärke

* Die hier aufgeführten Prozentwerte geben an, wie viele Befragten mit einer niedrigen/mittleren bzw. hohen Einstellungsstärke innerhalb einer Layoutgruppe die jeweilige Antwortkategorie für mindestens eines der sieben Items der PES bzw. eines der zehn Items des LIST-Inventars zur Antwortgabe gewählt haben.

Im ADD-DK-Layout besteht kaum ein Unterschied zwischen den beiden Einstellungsstärkegruppen. Hier wählen 37,8 % der Befragten mit einer niedrigen / mittleren und 39,7 % mit einer hohen Einstellungsstärke die rechte Extremkategorie. Im SEP-DK-Layout ist das Verhältnis schließlich umgekehrt. Mehr Befragte mit einer niedrigen / mittleren Einstellungsstärke (37,8 %) nutzen in

diesem Layout die untersuchte Antwortkategorie zur Beantwortung mindestens eines der PES-Items als Befragte mit einer hohen Einstellungsstärke (30,0 %).

Zwischen den Skalenlayouts zeigen sich für Befragte mit einer niedrigen / mittleren Einstellungsstärke folgende Unterschiede: Im NO-DK-Layout wird die visuell am äußeren Skalenrand dargestellte rechte Extremkategorie anteilig von den wenigsten Befragten (18,6 %) gewählt (vgl. Tabelle 6). In den beiden anderen Layouts, bei denen auf die rechte Extremkategorie noch die DK-Kategorie folgt, liegen die Befragtenanteile deutlich höher als im NO-DK-Layout, nämlich bei 38,1 % im SEP-DK- und bei 37,8 % im ADD-DK-Layout, und unterscheiden sich nicht untereinander. Auch unter den Befragten mit einer hohen Einstellungsstärke wird im NO-DK-Layout (25,0 %) anteilig von den wenigsten Befragten die rechte Extremkategorie gewählt. Die Rate liegt mit 39,7 % im ADD-DK-Layout deutlich höher und mit 30,0 % im SEP-DK-Layout dazwischen (vgl. Tabelle 6). Für beide Einstellungsgruppen werden die Unterschiede im Chi-Quadrat-Test als nicht signifikant unterschiedlich bewertet.

Zusammenfassend kann aus den Ergebnissen für die Analyse der PES-Items geschlossen werden, dass die Einstellungsstärke, wie in Hypothese (H 3) angenommen, moderierend auf den Einfluss des Layouts auf die Wahl der rechten Extremkategorie wirkt. Bei einer hohen Einstellungsstärke zeigen sich in deutlich geringerem Ausmaß Layouteffekte. Die Ergebnisse der Analyse der LIST-Items zeigen zwar ebenfalls Unterschiede in der Wahl der rechten Extremkategorie zwischen den Layoutgruppen, allerdings werden diese entgegen der Annahmen der Hypothese (H 3) nicht durch eine hohe Einstellungsstärke vermindert.

Mittelkategorie

Bei der Beantwortung der PES-Items wird in allen drei Layoutgruppen die Mittelkategorie anteilig von mehr Befragten mit einer niedrigen / mittleren Einstellungsstärke gewählt als von Befragten mit einer hohen Einstellungsstärke (vgl. Tabelle 6). Dies ist inhaltlich sinnvoll, denn Befragte mit einer hohen Einstellungsstärke sollten ihre Einstellung von Antwortkategorien links oder rechts der Skalenmitte repräsentiert sehen. Basierend auf dieser Annahme ist es erstaunlich, dass von Befragten mit einer hohen Einstellungsstärke 71,2 % im NO-

DK-, 64,4 % im SEP-DK- und noch 56,3 % im ADD-DK-Layout mindestens einmal die neutrale Mittelkategorie gewählt haben, um die Inhalte der PES-Items zu bewerten (vgl. Tabelle 6). Dieses Ergebnis steht jedoch in Einklang mit dem aktuellen Stand der Forschung in diesem Bereich, der besagt, dass Befragte nicht nur aus inhaltlichen, sondern auch aus anderen Gründen wie Ambivalenz gegenüber dem Inhalt, Satisficing oder Unklarheiten hinsichtlich der Fragestellung neutrale Antworten zu Items geben, zu denen sie eigentlich eine inhaltliche Meinung besitzen (vgl. z.B. Krosnick & Fabrigar, 1997; Schumann & Presser, 1981; Saris & Gallhofer 2007).

Unterschiede in der Wahl der Mittelkategorie sind sowohl für die Gruppe der Befragten mit einer niedrigen / mittleren Einstellungsstärke als auch für solche mit einer hohen Einstellungsstärke zwischen den drei Skalenlayouts zu sehen. Die höchste Rate an Befragten, die die Mittelkategorie wählen, ist jeweils im NO-DK-Layout zu beobachten, gefolgt vom SEP-DK- und schließlich ADD-DK-Layout. Die Unterschiede zwischen den Layoutgruppen sind bei Befragten mit einer niedrigen / mittleren Einstellungsstärke etwas größer als bei Befragten mit einer hohen ES, werden aber gemäß des Chi-Quadrat-Tests für beide Einstellungsstärkegruppen als nicht signifikant bewertet (vgl. Tabelle 6).

Bei Beantwortung der LIST-Items sind zwischen Befragten mit einer niedrigen / mittleren und einer hohen Einstellungsstärke keine Unterschiede in der Wahl der Mittelkategorie zu beobachten (vgl. Tabelle 6). Unabhängig von der Einstellungsstärke wählen etwa 10 Prozentpunkte mehr Befragte im NO-DK- und SEP-DK-Layout die Mittelkategorie als im ADD-DK-Layout; diese Unterschiede werden im Chi-Quadrat-Test als nicht signifikant bewertet.

Die Analyseergebnisse der Wahl der Mittelkategorie unter Berücksichtigung der Einstellungsstärke sprechen eher gegen die in Hypothese (H 3) getroffenen Annahmen. Bei den PES-Items ist die anteilig unterschiedliche Wahl der Mittelkategorie in den drei Layoutgruppen nicht nur für Befragte mit einer niedrigen / mittleren sondern auch für solche mit einer hohen Einstellungsstärke, wenn auch in etwas geringerem Maße, zu beobachten. Bei den LIST-Items zeigen sich keine Unterschiede zwischen den beiden Einstellungsstärkegruppen. Die Hypothese (H 1.2) über die Wahl der Mittelkategorie hingegen wird von den

Ergebnissen der Analyse der PES-Items unterstützt. Es scheint, dass im NO-DK-Layout die visuell eindeutig erkennbare Mittelkategorie die Befragten zu einer mittleren Antwortgabe verleitet. Die geringere Rate im SEP-DK-Layout, in dem ebenfalls eine visuell klar erkennbare Mittelkategorie vorliegt, wird - wie in Unterabschnitt (8.1) bereits diskutiert - auf die klar erkennbar vorgegebene DK-Kategorie zurückgeführt. Bei den LIST-Items hat das Angebot einer DK-Kategorie wiederum keinen Effekt auf die Wahl der Mittelkategorie, welche im NO-DK- und im SEP-DK-Layout gleichermaßen von den Befragten gewählt wird. Hier scheint die in Hypothese (H 1.3) getroffene Annahme bestätigt; der nach rechts verschobene visuelle Mittelpunkt und die dadurch nicht klar ersichtliche Mittelkategorie im ADD-DK-Layout scheinen zu einer geringeren Wahl dieser Antwortkategorie zu führen.

DK-Kategorie

Bei Beantwortung der PES-Items sind Unterschiede in der Wahl der DK-Kategorie zum einen zwischen Befragten mit einer niedrigen / mittleren und einer hohen Einstellungsstärke und zum anderen zwischen den beiden Skalenlayouts zu beobachten. Der größte signifikante Unterschied zwischen den beiden Layouts besteht in der Gruppe der Befragten mit einer niedrigen / mittleren Einstellungsstärke. Im SEP-DK-Layout wählen mit 30,6 % mehr als dreimal so viele Befragte die DK-Kategorie wie mit 8,7 % im ADD-DK-Layout. Werden nun die Befragten mit einer hohen Einstellungsstärke zwischen den beiden Layouts miteinander verglichen, so zeigt sich kein signifikanter Unterschied mehr (17,3 % vs. 16,7 %) (vgl. Tabelle 6). Dieses Ergebnis unterstützt die Hypothesen, dass die visuelle Hervorhebung der DK-Kategorie im SEP-DK-Layout Befragte anzieht (H 1.3) und zwar vor allem solche mit einer niedrigen / mittleren Einstellungsstärke (H 3).

Interessant ist der Unterschied in der Wahl der DK-Kategorie zwischen Befragten mit einer niedrigen / mittleren (8,7 %) und solchen mit einer hohen Einstellungsstärke (16,7 %) (vgl. Tabelle 6). Die höhere Rate an DK-Antworten unter Befragten mit einer hohen Einstellungsstärke erscheint inhaltlich nicht sinnvoll. Zwar wählen, wie bereits zuvor erwähnt, auch Befragte mit einer

relevanten Einstellung die DK-Kategorie zur Antwortgabe (vgl. z.B. Krosnick & Fabrigar, 1997; Schumann & Presser, 1981; Saris & Gallhofer 2007), da allerdings wie erwartet im SEP-DK-Layout Befragte mit einer niedrigen / mittleren Einstellungsstärke (30,6 %) deutlich häufiger die DK-Kategorie wählen als solche mit einer hohen Einstellungsstärke (17,3 %) (vgl. Tabelle 6), kann auch eine Fehlinterpretation der DK-Kategorie im ADD-DK-Layout die Ursache für den auffälligen Befund sein. Durch die nicht abgetrennte visuelle Darstellung der DK-Kategorie im ADD-DK-Layout ist es vorstellbar, dass Befragte die DK-Kategorie als inhaltliche Kategorie, nämlich als rechte Extremkategorie interpretieren und nutzen.

Bei den LIST-Items zeigen sich unerwartete, wenn auch nicht signifikante Unterschiede: Für die Gruppe der Befragten mit einer niedrigen / mittleren Einstellungsstärke wählen entgegen Annahme mehr Befragte im ADD-DK-Layout (5,4 %) die DK-Kategorie als Befragte im SEP-DK-Layout (2,4 %). Die erwartete höhere Rate an DK-Antworten im SEP-DK- im Vergleich zum ADD-DK-Layout zeigt sich hingegen in der Gruppe der Befragten mit einer hohen Einstellungsstärke (SEP-DK: 8,3 % vs. ADD-DK: 3,4 %) (vgl. Tabelle 6). Allerdings wurde für diese Gruppe angenommen, dass sie weniger Layouteffekte zeigen als Befragte mit einer niedrigen / mittleren Einstellungsstärke.

Insgesamt kann aus den Ergebnissen der PES-Items hinsichtlich der Wahl der DK-Kategorie der Schluss gezogen werden, dass wie in Hypothese (H 3) angenommen, die visuelle Darstellung der DK-Kategorie vor allem Befragte mit einer niedrigen / mittleren Einstellungsstärke in der Wahl bzw. Nicht-Wahl dieser Kategorie beeinflusst. Wird die DK-Kategorie mittels eines Trennstriches visuell hervorgehoben, zieht sie Befragte an. Wird die DK-Kategorie einfach an die inhaltlichen Kategorien angehängt, scheint sie von einem Teil der Befragten falsch interpretiert zu werden, nämlich als inhaltliche Antwortkategorie. Für die LIST-Items werden keine Effekte des Antwortskalenlayouts auf die Wahl der DK-Kategorie beobachtet. Dies unterstützt die Annahme der Hypothese (H 2), nach der Befragte bei Beantwortung der LIST-Items weniger empfänglich für das Layout der Skala sind als bei Beantwortung der PES-Items, weil sie die Fragen zu ihren

Lernstrategien einfacher aus dem Gedächtnis beantworten können als Fragen über abstraktere politische Inhalte (vgl. Unterabschnitt 5.3).

9. Ergebnisse der logistischen Multilevel-Regressionen

Dieser Abschnitt behandelt schließlich die Ergebnisse der logistischen Multilevel-Regressionen mit fixen Effekten. Die Vorstellung der Ergebnisse erfolgt dabei getrennt nach den drei abhängigen Variablen „Wahl der rechten Extremkategorie“ (Unterabschnitt 9.1), „Wahl der Mittelkategorie“ (Unterabschnitt 9.2) und „Wahl der DK-Kategorie“ (Unterabschnitt 9.3). Für jede dieser drei untersuchten Antwortkategorien werden die Ergebnisse der fünf, in Unterabschnitt (6.3.4) beschriebenen, logistischen Multilevel-Modelle erläutert und jeweils in einem Zwischenfazit im Hinblick auf die Hypothesen diskutiert.

9.1 Wahl der rechten Extremkategorie

Die zu erklärende Variable ist in den folgenden fünf logistischen Multilevel-Modellen die Wahl der rechten Extremkategorie. Diese Variable gibt für jedes der siebzehn untersuchten Items an, ob ein Befragter zu dessen Beantwortung die rechte Extremkategorie gewählt oder nicht gewählt hat. Somit wird die mittlere Neigung aller Befragten betrachtet, die rechte Extremkategorie zur Antwortabgabe zu nutzen, in Abhängigkeit der in den Modellen betrachteten erklärenden Variablen.

9.1.1 Modell 1: Effekt des Skalenlayouts

Modell 1 bezieht als Prädiktoren lediglich die nominale Variable „Skalenlayout“ mit ein, um die Effekte des Layouts auf die Wahl der rechten Extremkategorie zu testen. Das ADD-DK-Layout wird als Referenzkategorie gegenüber dem NO-DK- und dem SEP-DK-Layout definiert. Die beiden signifikanten, negativen Logit-Koeffizienten dieser nominalen Variable zeigen, dass, wie angenommen, die Wahrscheinlichkeit der Wahl der rechten Extremkategorie über alle untersuchten Items im NO-DK-Layout und im SEP-DK-Layout geringer ist als im ADD-DK-

Layout. Nach McFaddens- R^2 , ein Anpassungsmaß, welches den Grad der Verbesserung des vollständigen Modells, also des Modells mit erklärenden Variablen, gegenüber dem Nullmodell, also dem Modell ohne erklärenden Variablen, widerspiegelt, von 0,078 kommt es durch die Berücksichtigung des Skalenlayouts im Modell zu einer Steigerung dessen Erklärungskraft im Vergleich zu einem Null-Modell, also ein Modell ohne Prädiktoren.¹¹

Wie zuvor in Unterabschnitt (6.3.3) erläutert, können Logit-Koeffizienten untereinander zwar in ihrer Richtung verglichen und in ihrer Signifikanz beurteilt werden, nicht aber hinsichtlich ihrer Stärke und ihrer inhaltlichen Relevanz. Um eine Aussage über die Stärke der Effekte der einzelnen Skalenlayouts auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie treffen zu können, werden im Folgenden die durchschnittlichen marginalen Effekte (auch als Effekt auf die Wahrscheinlichkeit bezeichnet, kurz AME) sowie die vorhergesagten Wahrscheinlichkeiten zur Ergebnisinterpretation herangezogen.

Der Graph in Abbildung 10 stellt die AMEs inklusive deren 95 %-Konfidenzintervalle des NO-DK-Layouts und des SEP-DK-Layouts in Bezug zu dem AME der Referenzgruppe „ADD-DK-Layout“ dar, welcher im Graphen durch die Nulllinie repräsentiert wird. Wie in Unterabschnitt (6.3.3) beschrieben, sind die AMEs nominaler Variablen einfach als mittlere Differenz zwischen den vorhergesagten Wahrscheinlichkeiten einer Ausprägung einer nominalen Variable (z.B. NO-DK-Layout) in Bezug zur Referenzkategorie (hier ADD-DK-Layout) zu interpretieren. Die AMEs im Graphen in Abbildung 10 zeigen, dass der Effekt auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie im NO-DK-Layout 4 Prozentpunkte und im SEP-DK-Layout 2 Prozentpunkte signifikant schwächer ist als der Effekt des ADD-DK-Layouts (vgl. auch Tabelle 8, Seite 99).

Nachdem die mittleren Unterschiede in der Wahl der rechten Extremkategorie anhand der AMEs betrachtet wurden, ist noch die absolute Höhe der vorhergesagten Wahrscheinlichkeiten einer Antwortabgabe anhand der rechten Extremkategorie in den drei Skalenlayouts von Interesse.

¹¹ Der Wertebereich von McFaddens- R^2 reicht von Null bis unter Eins, wobei höhere Werte einer besseren Anpassung des Modells an die Daten und ein Wert von Null einem Modell ohne Erklärungskraft entsprechen.

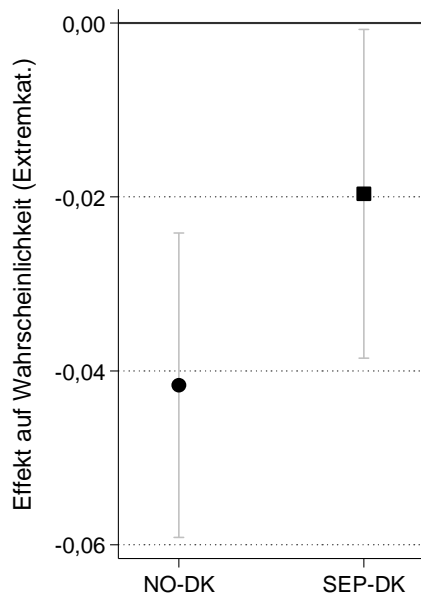


Abbildung 10 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkat. nach Layout (Referenz: ADD-DK)

Im Graphen in Abbildung 11 ist zu sehen, dass Befragte im ADD-DK-Layout mit 9,7 % eine deutlich höhere vorhergesagte Wahrscheinlichkeit für eine extreme Antwort besitzen als Befragte im NO-DK-Layout mit 5,5 %. Die Gruppe im SEP-DK-Layout liegt mit einer vorhergesagten Wahrscheinlichkeit von 7,7 % dazwischen (vgl. auch Tabelle 8, Seite 99).

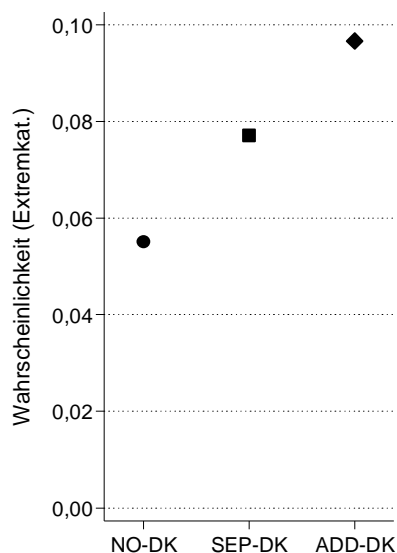


Abbildung 11 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkat. nach Layout

9.1.2 Modell 2: Effekt des Item-Konstrukts

In Modell 2 wird zusätzlich zum Skalenlayout die nominale Variable „Item-Konstrukt“ eingeführt. Mittels dieses Prädiktors werden die untersuchten Items den beiden Konstrukten „PES“ und „LIST“ zugeteilt. Somit kann untersucht werden, ob Befragte für die beiden Item-Konstrukte unterschiedliche Neigungen in der Wahl der rechten Extremkategorie besitzen. Der Logit-Koeffizient für das LIST-Konstrukt (mit dem PES-Konstrukt als Referenzkategorie) nimmt einen signifikanten, negativen Wert an. Danach ist die mittlere Neigung aller Befragten, die rechte Extremkategorie zu wählen für die LIST-Items geringer als bei den PES-Items. Auch die AMEs weisen auf signifikant niedrigeren Effekt auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie bei den LIST-Items hin. Der Unterschied beträgt 7,7 Prozentpunkte (vgl. auch Tabelle 8, Seite 99).

Bei der Beantwortung der PES-Items haben Befragte mit 12,9 % eine signifikant höhere vorhergesagte Wahrscheinlichkeit, die rechte Extremkategorie zu wählen, als mit 5,2 % bei Beantwortung der LIST-Items (vgl. Tabelle 9, Seite 100).

Auch unter Berücksichtigung der nominalen Variable „Item-Konstrukt“ in Modell 2 bleiben dessen weitere Ergebnisse im Vergleich zum vorangehenden Modell 1 unverändert: Unter Kontrolle der jeweiligen Zugehörigkeit der Items zu den beiden Item-Konstrukten PES bzw. LIST ist die Wahrscheinlichkeit der Wahl der rechten Extremkategorie im NO-DK-Layout und im SEP-DK-Layout signifikant niedriger als im ADD-DK-Layout. Der leichte Anstieg von McFaddens- R^2 von 0,078 (Modell 1) auf 0,083 (Modell 2) und die Verringerungen des AIC und des BIC sprechen zudem dafür, dass die Hinzunahme der Variable „Item-Konstrukt“ die Erklärungskraft des Modells verbessert (vgl. Tabelle 7, Seite 98).¹²

¹² Mit steigender Zahl an erklärenden Variablen im Modell nimmt auch McFaddens- R^2 prinzipiell größere Werte an. Da es aber Ziel ist, Modelle zu entwickeln, die zum einen gut an die Daten angepasst sind, zum anderen aber auch möglichst sparsam formuliert sind, sollte die Güte des Modells nicht alleine anhand von McFaddens- R^2 beurteilt werden. Deshalb werden hier auch noch das „Akaike Informationskriterium“ (kurz AIC) und das „Bayessche Informationskriterium“ (kurz BIC) herangezogen, denn sie berücksichtigen die Zahl der Modellparameter und erlauben zudem, sowohl geschachtelte als auch nicht-geschachtelte Modelle miteinander zu vergleichen. Sowohl das AIC als auch das BIC haben einen Wertebereich von Null bis plus. Unendliche, kleine Werte weisen auf ein besser angepasstes Modell hin (Best und Wolf 2010).

9.1.3 Modell 3: Interaktionseffekt „Item-Konstrukt × Layout“

Modell 3 stellt eine Erweiterung von Modell 2 um den Interaktionsterm „Konstrukt × Layout“ dar. Durch diese Interaktion wird es möglich zu beurteilen, ob, wie in Hypothese (H 2) angenommen, Befragte bei der Beantwortung der Item-Konstrukte nicht gleichermaßen stark durch das Layout in der Wahl der rechten Extremkatgorie beeinflusst werden (vgl. Unterabschnitte 4.3 und 5.3). Die beiden positiven Logit-Koeffizienten der Interaktion sind nicht signifikant. Allerdings ist die Interpretation von Interaktionen in logistischen Regressionen, wie in Unterabschnitt (6.3.3) beschrieben, direkt aus der Ergebnistabelle schwierig, weshalb im Folgenden eine Interpretation auf Basis der graphischen Darstellung der AMEs vorgenommen wird. Aus den beiden Graphen in Abbildung 12 wird deutlich, dass die Effekte auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie für die PES-Items deutlich stärker sind als für die LIST-Items.

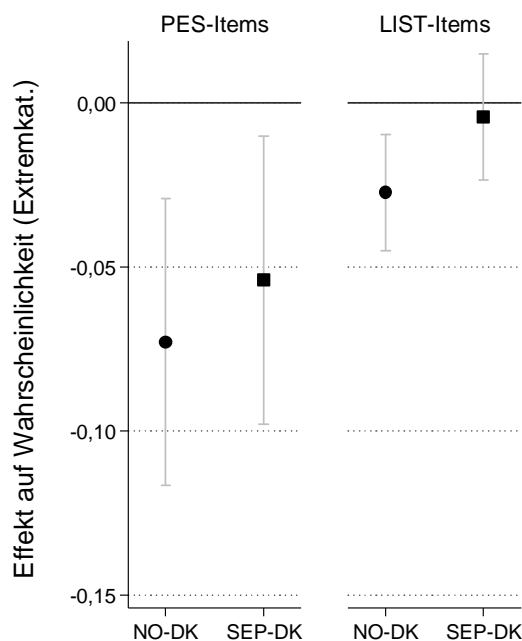


Abbildung 12 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Item-Konstrukt

Für die PES-Items werden sowohl im NO-DK-Layout (AME= -0,073) als auch im SEP-DK-Layout (AME= -0,054) signifikante, negative Effekte auf die Wahrscheinlichkeit einer extremen Antwortabgabe im Vergleich zum ADD-DK-Layout ausgegeben. Für die LIST-Items unterscheiden sich lediglich die Effekte auf

die Wahrscheinlichkeit des NO-DK-Layouts und des ADD-DK-Layouts signifikant um 3 Prozentpunkte voneinander. Damit ist dieser Effektunterschied nicht einmal halb so groß wie der Effektunterschied der selben Layoutgruppen bei Beantwortung der PES-Items (7,3 Prozentpunkte).

Die vorhergesagten Wahrscheinlichkeiten der Wahl der rechten Extremkategorie, dargestellt in den Graphen in Abbildung 13, sind für die PES-Items in allen drei Layoutgruppen höher als für die LIST-Items. Zudem sind die Unterschiede zwischen den Layoutgruppen bei den PES-Items größer: Hier haben Befragte im ADD-DK-Layout mit 17,3 % die höchste vorhergesagte Wahrscheinlichkeit, die rechte Extremkategorie zu wählen, Befragte im SEP-DK-Layout hingegen nur eine vorhergesagte Wahrscheinlichkeit von 11,9 % und im NO-DK-Layout schließlich von 10,0 %. Bei den LIST-Items liegen die vorhergesagten Wahrscheinlichkeiten in den drei Layoutgruppen zwischen 3,6 % und 6,3 %.

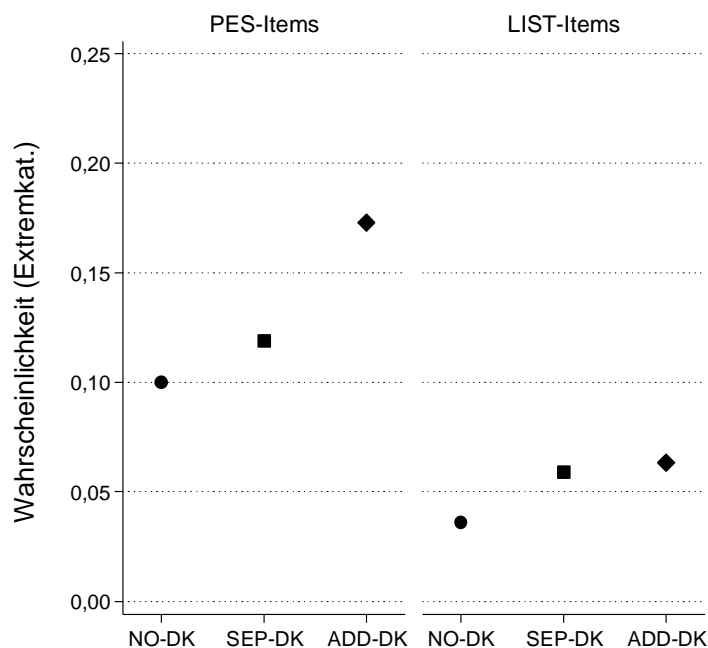


Abbildung 13 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout und Item-Konstrukt

Die weiteren Ergebnisse der logistischen Multilevel-Regression von Modell 3 sind unverändert im Vergleich zu denen von Modell 2 (vgl. Tabelle 7, Seite 98): Auch unter Berücksichtigung der Interaktion „Item-Konstrukt \times Layout“ bleibt die höhere Wahrscheinlichkeit einer Antwortabgabe anhand der rechten Extremkategorie in der

Gruppe des ADD-DK-Layouts gegenüber den anderen beiden Layoutgruppen bestehen. Genauso zeigt sich weiterhin eine höhere Wahrscheinlichkeit für die Wahl der rechten Extremkategorie bei Beantwortung der PES- im Vergleich zu den LIST-Items. Allerdings trägt die Berücksichtigung der Interaktion „Item-Konstrukt \times Layout“ nicht zu einer Verbesserung der Erklärungskraft des Modells bei. Weder steigt McFaddens- R^2 noch verringern sich die AIC- und BIC-Werte.

9.1.4 Modell 4: Effekt der Einstellungsstärke

Modell 4 integriert zusätzlich noch die metrische Variable „Einstellungsstärke“ gegenüber den thematischen Inhalten der beiden Konstrukte „PES“ und „LIST“ in die logistische Multilevel-Regression. Somit kann der in Hypothese (H 3) angenommene Zusammenhang zwischen der individuellen Einstellungsstärke und der Wahrscheinlichkeit einer Antwortabgabe mittels der rechten Extremkategorie betrachtet werden.

In der Ergebnistabelle wird ein signifikanter, positiver Logit-Koeffizient für die Einstellungsstärke ausgewiesen. Demnach steigt die Wahrscheinlichkeit einer extremen Antwortabgabe über alle untersuchten Items mit zunehmender Einstellungsstärke. Gemäß dem AME für die Einstellungsstärke steigt die Wahrscheinlichkeit der Wahl der rechten Extremkategorie um 0,3 Prozentpunkte, wenn sich die Einstellungsstärke um eine Einheit erhöht (vgl. Tabelle 8, Seite 99).

Dieser positive Zusammenhang zwischen Einstellungsstärke und Wahrscheinlichkeit der Wahl der rechten Extremkategorie zeigt sich auch deutlich in den vorhergesagten Wahrscheinlichkeiten, dargestellt in dem Graphen in Abbildung 14: Wählen lediglich 4 % der Befragten mit einer sehr niedrigen Einstellungsstärke von 1 Punkt die rechte Extremkategorie zur Antwortabgabe, verdreifacht sich die Wahrscheinlichkeit auf 12 % für Befragte mit einer sehr hohen Einstellungsstärke von 25 Punkten.

Die inhaltliche Interpretation des positiven Zusammenhangs zwischen Einstellungsstärke und der Wahrscheinlichkeit der Wahl der rechten Extremkategorie wurde bereits in der bivariaten Analyse beobachtet und an dieser Stelle diskutiert (vgl. Unterabschnitt 8.2).

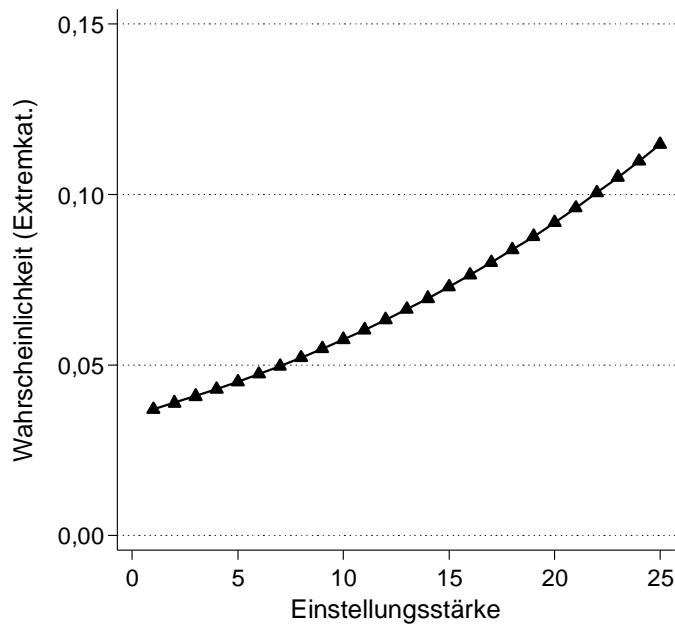


Abbildung 14 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Einstellungsstärke

Abschließend werden auch für Modell 4 noch die weiteren Ergebnisse der logistischen Multilevel-Regression betrachtet: Hierbei zeigt wieder, dass die Ergebnisse im Vergleich zum vorangehenden Modell 3 durch die Hinzunahme der Variable „Einstellungsstärke“ unbeeinflusst bleiben. Demnach liegt auch unter Kontrolle der Einstellungsstärke die höchste Wahrscheinlichkeit einer Antwortabgabe anhand der rechten Extremkategorie in der Gruppe des ADD-DK-Layouts gefolgt vom SEP-DK- und schließlich NO-DK-Layout. Genauso zeigt sich weiterhin eine höhere Wahrscheinlichkeit für die extreme Antwortabgabe bei Beantwortung der PES- im Vergleich zu den LIST-Items. Des Weiteren ist ein Anstieg von McFaddens- R^2 von 0,084 (Modell 3) auf 0,092 (Modell 4) zu beobachten. Daraus lässt sich schließen, dass die Erklärungskraft dieses Modells durch die Hinzunahme des weiteren Prädiktors steigt. Auch die Verringerung der Werte des AIC und des BIC gegenüber den Werten des vorangehenden Modells 3 spricht dafür, dass durch die Modellerweiterung dessen Güte angehoben wird (vgl. Tabelle 7, Seite 98).

9.1.5 Modell 5: Interaktionseffekt „Einstellungsstärke × Layout“

Im letzten Modell 5 wird der Interaktionsterm „Einstellungsstärke × Layout“ eingeführt. Damit kann die Annahme der Hypothese (H 3) überprüft werden, dass mit steigender Einstellungsstärke die Effekte des Layouts ihre Signifikanz verlieren.

In der Ergebnistabelle des logistischen Multilevel-Modells wird für die Interaktion „Einstellungsstärke × NO-DK-Layout“ ein signifikanter, positiver Logit-Koeffizient ausgewiesen. Der Koeffizient der Interaktion „Einstellungsstärke × SEP-DK-Layout“ ist hingegen nicht signifikant. Für eine einfachere inhaltliche Interpretation dieser Interaktion wird im Folgenden wieder die graphische Darstellung der AMEs bzw. der vorhergesagten Wahrscheinlichkeiten genutzt. Die AMEs (vgl. Abbildung 15) verdeutlichen, dass die Effekte auf die Wahrscheinlichkeit einer Antwortabgabe anhand der rechten Extremkategorie im Bereich der unteren bis mittleren Einstellungsstärke im NO-DK-Layout signifikant niedriger ausfallen als in der Referenzkategorie „ADD-DK-Layout“. Im Bereich der höheren Einstellungsstärke gehen die Effekte wie angenommen gegen Null und unterscheiden sich demnach nicht mehr voneinander. Zwischen dem SEP-DK- und dem ADD-DK-Layout wird hingegen nur für einen engen Bereich der mittleren Einstellungsstärke ein signifikanter Unterschied der Effekte auf die Wahrscheinlichkeit angezeigt.

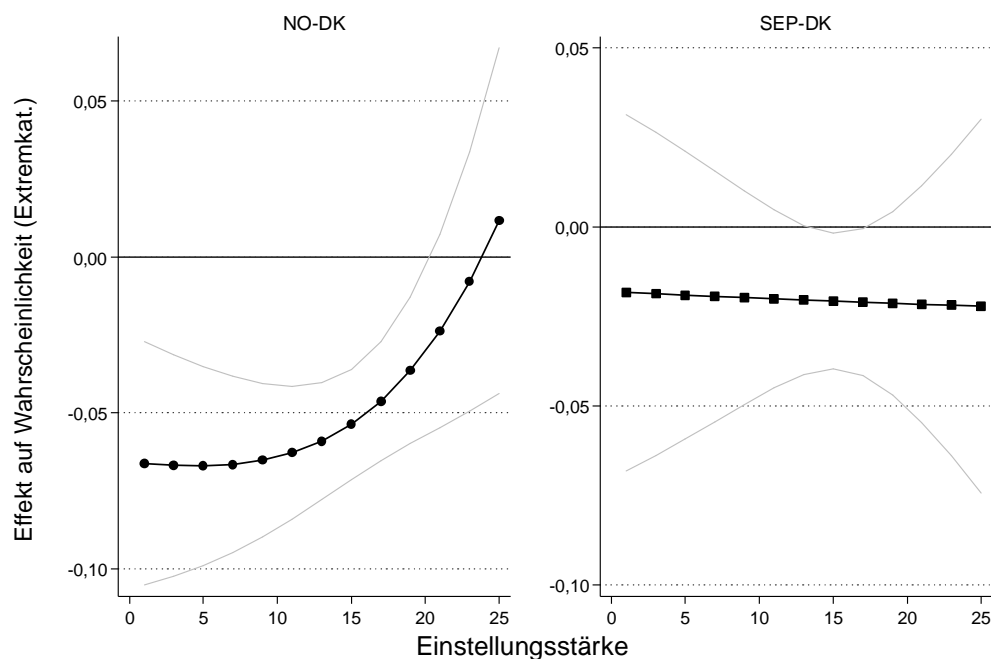


Abbildung 15 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Einstellungsstärke

Die vorhergesagten Wahrscheinlichkeiten der einzelnen Ausprägungen der Einstellungsstärke nach Skalenlayout, dargestellt in den drei Graphen in Abbildung 16, unterscheiden sich im NO-DK-Layout auf der einen und SEP-DK- und ADD-DK-Layout auf der anderen Seite: Die Wahrscheinlichkeit der Wahl der rechten Extremkategorie liegt für Befragte mit einer Einstellungsstärke im unteren Bereich (Wert zwischen 1 bis 10 Punkte) im NO-DK-Layout mit 0,8 % bis 2,4 % deutlich niedriger als im SEP-DK-Layout mit 5,6 % bis 6,8 % und im ADD-DK-Layout mit 7,4 % bis 8,8 %. Bis zu einer mittleren Einstellungsstärke von 15 Punkten haben Befragte im NO-DK-Layout eine deutlich geringere Wahrscheinlichkeit einer extremen Antwortabgabe im Vergleich zu Befragten mit der niedrigsten Einstellungsstärke in den beiden anderen Layouts. Aufgrund des starken Anstiegs der Wahrscheinlichkeiten im Bereich der mittleren und hohen Einstellungsstärke im NO-DK-Layout ändert sich das Bild jedoch im Bereich der höheren Einstellungsstärke ab 18 Punkten. Hier bestehen keine größeren Unterschiede mehr zwischen den drei Layoutgruppen. Die vorhergesagten Wahrscheinlichkeiten der extremen Antwortabgabe liegen auf einem ähnlichen Niveau. Befragte im NO-DK-Layout haben bei der stärksten Einstellungsstärke von 25 Punkten nun eine vorhergesagte Wahrscheinlichkeit von 12,8 %, die rechte Extremkategorie zu wählen, im SEP-DK-Layout von 9,4 % und im ADD-DK-Layout 11,7 %.

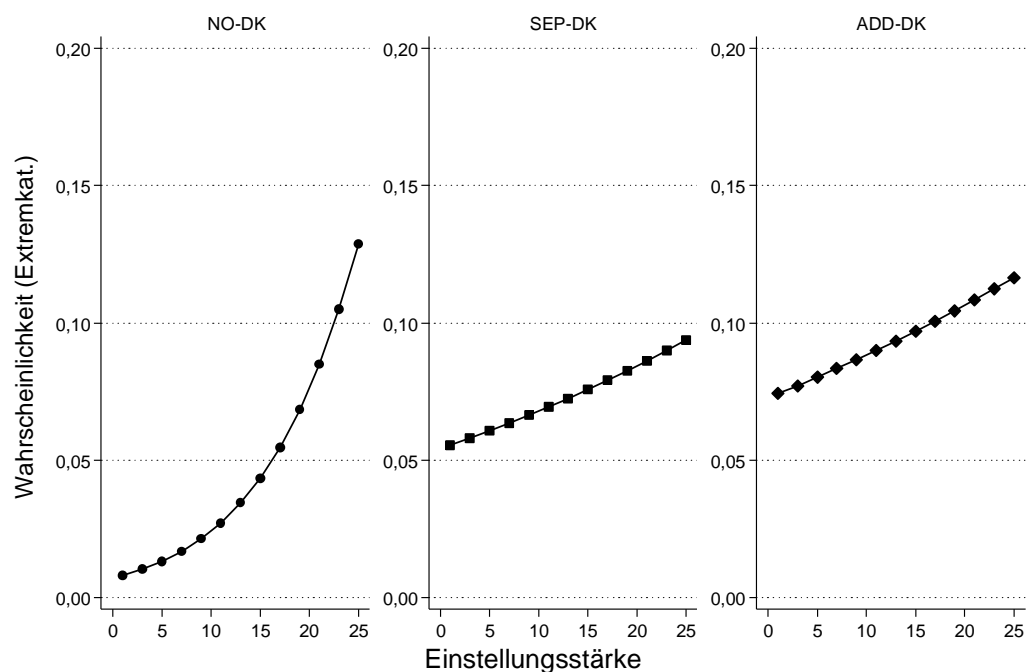


Abbildung 16 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout und Einstellungsstärke

Diese Ergebnisse lassen den Schluss zu, dass Befragte mit einer niedrigen bis mittleren Einstellungsstärke die Antwortskala im NO-DK-Layout hinsichtlich der Wahl der rechten Extremkategorie anscheinend anders zur Antwortabgabe nutzen als Befragte in einem Layout mit einer DK-Kategorie. Der Verlauf der Kurve der vorhergesagten Wahrscheinlichkeiten der Befragten im NO-DK-Layout kommt dabei der theoretischen Annahme über das Antwortverhalten am nächsten: Befragte, die keine starke Einstellung zu einem abgefragten Thema besitzen, sollten auch eher keine Extremkategorie nutzen, um das Item damit zu beantworten. Eine Erklärung, warum Befragte in den beiden Layouts mit DK-Kategorie über alle Ausprägungen der Einstellungsstärke hinweg eine höhere Chance haben, die rechte Extremkategorie zu wählen, könnte lauten, dass durch die sechsstufige Erscheinung der Skala aufgrund der vorgegebenen DK-Kategorie die Extremkategorie nicht mehr als extrem wahrgenommen wird wie im NO-DK-Layout und daher auch von Befragten mit einer mittleren Einstellungsstärke gewählt wird.

Aus den Ergebnissen der vorangehenden Modelle 3 und 4 ist bekannt, dass Befragte die rechte Extremkategorie für die PES- und LIST-Items mit sehr unterschiedlichen Wahrscheinlichkeiten wählen. Daher wird im Folgenden der Einfluss der Einstellungsstärke auf das Auftreten von Layouteffekten nochmals getrennt nach Item-Konstrukt betrachtet.

Die vier Graphen in Abbildung 17 veranschaulichen die AMEs der einzelnen Stufen der Einstellungsstärke auf die Wahrscheinlichkeit einer extremen Antwortabgabe getrennt nach Skalenlayout und Item-Konstrukt. Die beiden Graphen auf der linken Seite zeigen für das NO-DK-Layout, dass für Befragte mit einer niedrigen bis mittleren Einstellungsstärke, sowohl bei den PES-Items (Graph oben links) als auch bei den LIST-Items (Graph unten links), signifikant niedrigere Effekte auf die Wahrscheinlichkeit einer extremen Antwortabgabe ausgewiesen werden als für Befragte der Referenzgruppe „ADD-DK-Layout“. Der Unterschied zwischen den Effekten der beiden Layouts ist für die PES-Items zwar deutlich größer als für die LIST-Items, allerdings gleichförmig: Für einen Befragten mit einer mittleren Einstellungsstärke von 10 Punkten wird ein Effektunterschied bei den PES-Items von 11,2 Prozentpunkten, bei den LIST-Items hingegen von nur 4,2 Prozentpunkten im Vergleich zum ADD-DK-Layout berechnet. Bei beiden Item-Konstrukten

verringert sich der Effektunterschied mit steigender Einstellungsstärke bis er im oberen Bereich ab einer Einstellungsstärke von 20 Punkten nicht mehr signifikant ist. Für die AMEs ab einer Einstellungsstärke von 21 Punkten zeigen die Nulllinie schneidenden Konfidenzintervalle, dass zwischen dem Effekt des NO-DK-Layouts und dem des ADD-DK-Layouts kein signifikanter Unterschied mehr besteht.

Die beiden Graphen auf der rechten Seite in Abbildung 17 zeigen die Ergebnisse für das SEP-DK-Layout in Bezug zum ADD-DK-Layout, wiederum nach Stärke der Einstellung und nach Item-Konstrukt. Generell werden für den Vergleich dieser Layoutgruppen sehr geringe Effekte für beide Item-Konstrukte ausgegeben: Lediglich im Bereich einer mittleren Einstellungsstärke unterscheiden sich die Effekte auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie für die PES-Items signifikant voneinander.

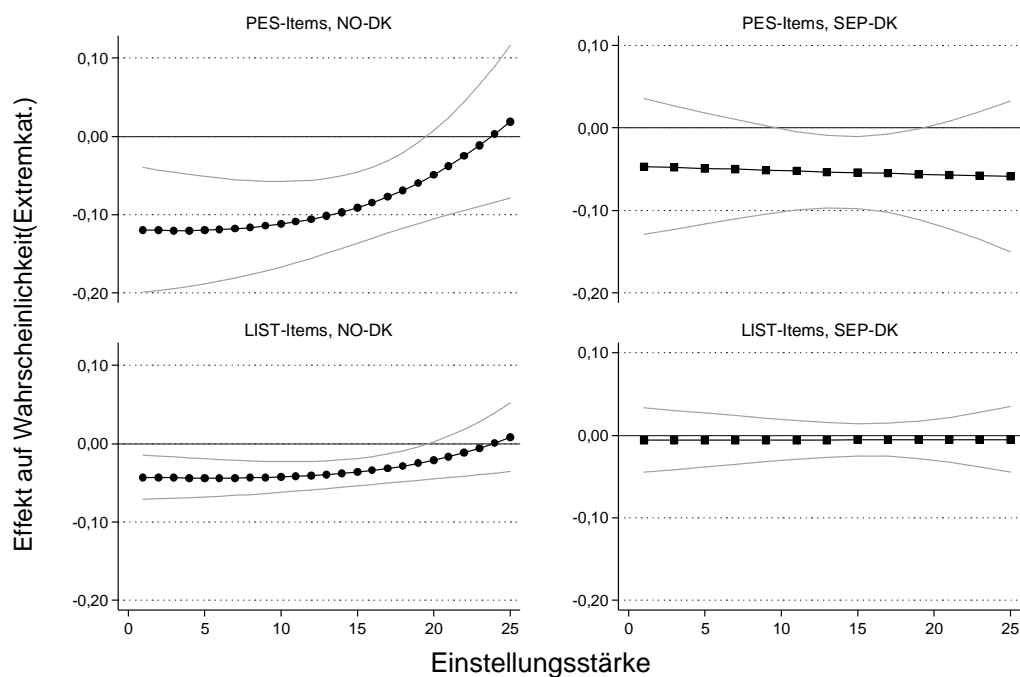


Abbildung 17 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK), Einstellungsstärke und Item-Konstrukt

Schließlich zeigen die Wahrscheinlichkeiten für die einzelnen Stufen der Einstellungsstärke, getrennt betrachtet nach Layoutgruppen und nach Item-Konstrukt (vgl. Abbildung 18), einen sehr ähnlichen Verlauf wie in der zuvor beschriebenen gemeinsamen Betrachtung aller untersuchten Items (vgl. Abbildung 16). Für beide Item-Konstrukte ist zu sehen, dass für das NO-DK-Layout im

unteren Bereich der Einstellungsstärke eine sehr niedrige Wahrscheinlichkeit für eine Antwortabgabe anhand der rechten Extremkatgeorie vorhergesagt wird. Diese steigt mit zunehmender Einstellungsstärke exponentiell an. Im Vergleich dazu wird für die beiden anderen Layoutgruppen für die niedrige Einstellungsstärke eine höhere Wahrscheinlichkeit für die Wahl der rechten Extremkatgorie ausgewiesen. Diese Wahrscheinlichkeit steigt mit zunehmender Einstellungsstärke moderat und linear an. Im oberen Bereich der Einstellungsstärke unterscheiden sich die drei Layoutgruppen in beiden Konstrukten nicht mehr wesentlich voneinander. Insgesamt liegen die vorhergesagten Wahrscheinlichkeiten für die PES-Items höher als für die LIST-Items, was aus den bisherigen Ergebnissen bereits bekannt ist.

Der Modellvergleich zeigt, dass der Logit-Koeffizient des SEP-DK-Layouts in Modell 5 unter Berücksichtigung der Interaktion „Einstellungsstärke \times Layout“ seine Signifikanz verliert. Demnach wählen Befragte im SEP-DK-Layout und im ADD-DK-Layout gleichermaßen wahrscheinlich die rechte Extremkatgorie zur Antwortabgabe, wenn ihre individuelle Einstellungsstärke gegenüber den Inhalten der Item-Konstrukte kontrolliert wird. Die weiteren Prädiktoren des Modells 5 bleiben in ihren Effekten im Vergleich zu Modell 4 unbeeinflusst. Unter Hinzunahme der Interaktion „Einstellungsstärke \times Layout“ steigt McFaddens- R^2 von 0,092 (Modell 4) auf 0,098 (Modell 5) an. Zudem verringern sich die Werte des AIC und des BIC (vgl. Tabelle 7, Seite 98). Daher scheint die Einbeziehung dieser Interaktion die Erklärungskraft des Modells anzuheben.

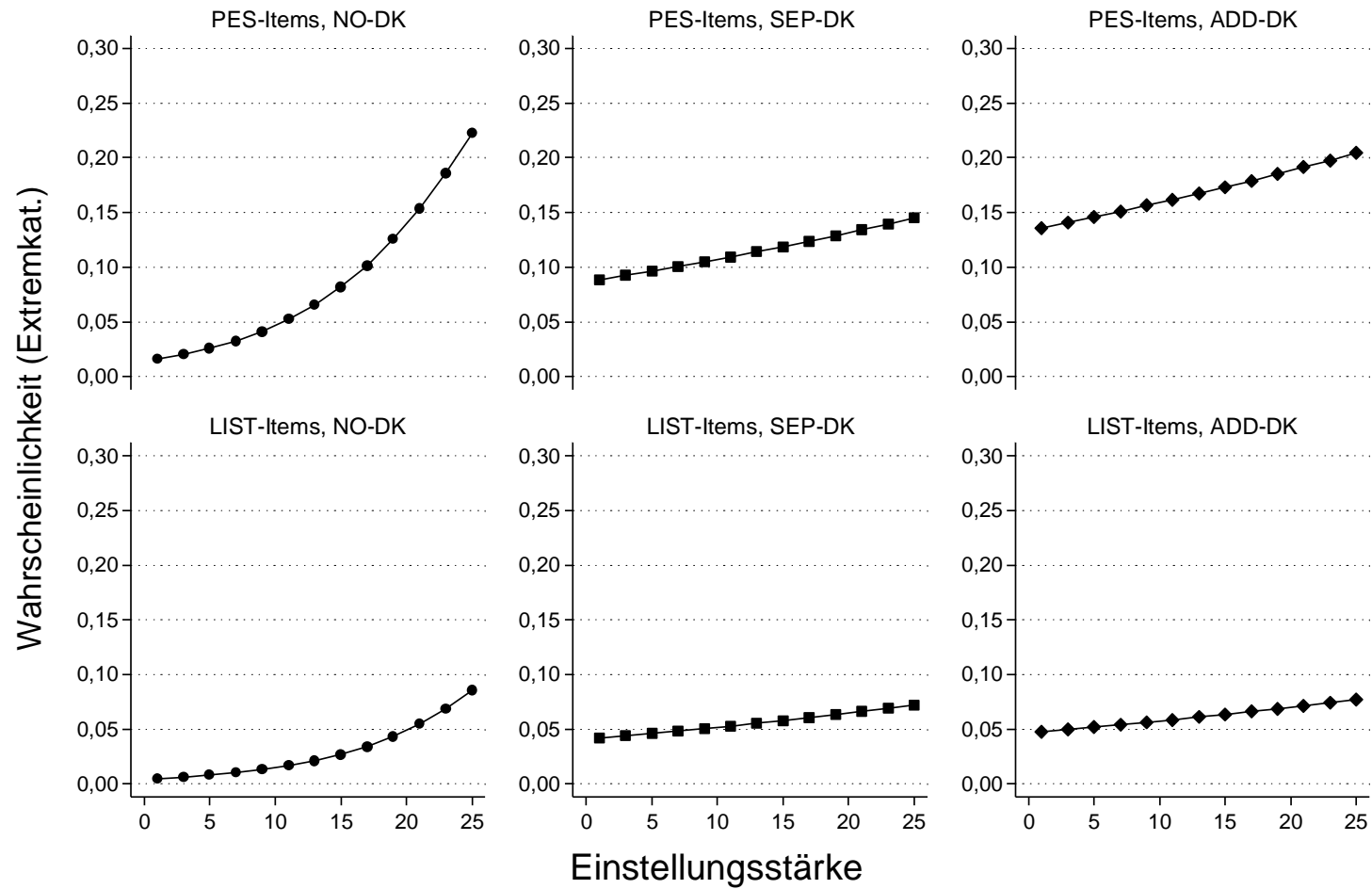


Abbildung 18 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkategorie nach Layout, Einstellungsstärke und Item-Konstrukt

9.1.6 Zwischenfazit

Effekt des Layouts der Ratingskala

Die Hypothese (H 1.1) besagt hinsichtlich des Einflusses des Skalenlayouts, dass Befragte im ADD-DK-Layout die höchste Wahrscheinlichkeit für eine Antwortabgabe anhand der rechten Extremkategorie besitzen. Begründet wird dies durch die Annahme, dass die Antworten der Befragten durch den verschobenen visuellen Skalenmittelpunkt im ADD-DK-Layout hin zum rechten Skalenende gezogen werden. Zudem wirkt dieses Skalenlayout durch die angefügte DK-Kategorie visuell breiter, nämlich sechsstufig, was ebenfalls eine Verschiebung der Antworten zum rechten Skalenende hin verursachen sollte. Dahingegen wird für Befragte im NO-DK-Layout die geringste Wahrscheinlichkeit der Wahl der rechten Extremkategorie erwartet, denn in diesem Layout stimmt der inhaltliche mit dem visuellen Mittelpunkt überein. Zudem ist das rechte, extreme Skalenende klar erkennbar. Daher sollten Befragte nicht in ihrer Wahrnehmung der Skala fehlgeleitet werden. Das SEP-DK-Layout stellt schließlich eine Mischung der beiden zuvor genannten Layouts dar: Einerseits erscheint die Skala durch das Vorhandensein einer DK-Kategorie visuell breiter als die Skala im NO-DK-Layout, nämlich sechststufig. Andererseits stimmen der inhaltliche und visuelle Mittelpunkt weiterhin wegen der Abtrennung der DK-Kategorie mittels eines Trennstrichs überein. Die DK-Kategorie sollte für den Befragten also klar erkennbar, nämlich als inhaltlich nicht zur Ratingskala zugehörig, wahrgenommen werden. Daher wird angenommen, dass Befragte in diesem Layout ebenfalls eine niedrigere Wahrscheinlichkeit für eine extreme Antwortabgabe besitzen als Befragte im ADD-DK-Layout (vgl. Unterabschnitt 5.3).

Wie angenommen zeigen die Ergebnisse, dass das Skalenlayout als erklärende Variable die vorhergesagten Wahrscheinlichkeiten der Wahl der rechten Extremkategorie signifikant beeinflusst (vgl. Modell 1). Am höchsten ist die Wahrscheinlichkeit im ADD-DK- und am niedrigsten im NO-DK-Layout; die des SEP-DK-Layouts liegt dazwischen. Daher ist zu schließen, dass die Verschiebung des visuellen Mittelpunktes und die sechsstufige visuelle Erscheinung im ADD-DK-Layout die Skala für Befragte am rechten Skalenende breiter erscheinen und sie wahrscheinlicher die rechte

Extremkategorie wählen lässt. Im Vergleich dazu wählen Befragte im NO-DK-Layout mit klarem fünfstufigem Aufbau und mit übereinstimmenden visuellen und inhaltlichen Mittelpunkt und im SEP-DK-Layout mit klar erkennbarem Mittelpunkt die rechte Extremkategorie weniger wahrscheinlich.

Effekt des Item-Konstrukts

In Hypothese (H 2) wird angenommen, dass die Ergebnisse über die Effekte des Skalenlayouts auf die Wahl der rechten Extremkategorie unterschiedlich für die beiden Item-Konstrukte „PES“ und „LIST“ ausfallen. Die Items der PES erfassen das Konzept der politischen Selbstwirksamkeit. Darin beurteilen Befragte, inwiefern sie selbst und die Bürger Einfluss auf das politische Geschehen nehmen können. Hier lautet die Annahme, dass dieses Thema für die meisten Befragten eher abstrakt ist und sie keine vorgefertigte, aus dem Gedächtnis einfach abrufbereite Meinung dazu besitzen, die sie nur wiedergeben müssten. Im Gegensatz dazu stehen die Items des LIST-Inventars. Diese Items fragen das Zutreffen bestimmter Aussagen über Lernsituationen ab, welche die befragten Studenten aus ihrem Studienalltag kennen sollten. Daher ist hier die Annahme, dass die Befragten bei der Beantwortung der PES-Items mehr kognitive Leistung vollbringen müssen als bei den LIST-Items, was sie empfänglicher für den Fragekontext und somit für das Antwortskalenlayout werden lassen sollte (vgl. Unterabschnitt 5.3).

Die Betrachtung der Wahl der rechten Extremkategorie, getrennt nach den beiden untersuchten Item-Konstrukten „PES“ und „LIST“, zeigt, dass Befragte mit einer deutlich höheren vorhergesagten Wahrscheinlichkeit die rechte Extremkategorie bei Beantwortung der PES-Items im Vergleich zu den LIST-Items wählen (vgl. Modell 2 in Unterabschnitt 9.1.2).

Offen ist die Frage, ob der Unterschied in der Wahl der rechten Extremkategorie zwischen den Item-Konstrukten inhaltlich begründet ist, also ob Befragte tatsächlich eine deutlichere Zustimmung zu den Inhalten der PES als zu denen des LIST-Inventars besitzen. Oder ob, wie in Hypothese (H 2) formuliert, bei den abstrakteren PES-Items die Layouteffekte die Befragten in ihrer Antwortgabe mehr beeinflussen als bei den

LIST-Items. Um diese Frage beantworten zu können, wurde eine Interaktion von Item-Konstrukt und Antwortskalenlayout aufgenommen (vgl. Modell 3 in Unterabschnitt 9.1.3). Hier zeigen die Ergebnisse, dass die vorhergesagte Wahrscheinlichkeit der Wahl der rechten Extremkategorie bei den PES-Items hier ebenfalls insgesamt auf einem deutlich höheren Niveau liegt als bei den LIST-Items. Die Effekte auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie sind ebenfalls für die PES-Items größer, was sich bei diesen Items in größeren Unterschieden zwischen den drei Layoutgruppen hinsichtlich ihrer Wahrscheinlichkeit einer extremen Antwortabgabe widerspiegelt. Daraus kann geschlossen werden, dass Befragte bei Beantwortung der PES-Items zum einen den Inhalten generell stärker zustimmen als denen der LIST-Items und zum anderen in ihrer Antwortgabe auch stärker vom Layout, vor allem dem ADD-DK-Layout, hin zum rechten Skalenende beeinflusst werden.

Effekt der Einstellungsstärke

Für den Einfluss der Einstellungsstärke von Befragten auf ihre Empfänglichkeit für das Antwortskalenlayout wird in Hypothese (H 3) angenommen, dass Befragte mit steigender Einstellungsstärke weniger durch den Fragekontext beeinflusst werden und somit die Layouteffekte ihre Signifikanz verlieren und keine Unterschiede in der Wahrscheinlichkeit der Wahl der rechten Extremkategorie zwischen den Layoutgruppen mehr beobachtet werden können (vgl. Unterabschnitt 5.3).

Die Berücksichtigung der Einstellungsstärke gegenüber der politischen Selbstwirksamkeit und den persönlichen Lernstrategien im Studium zeigt, dass die vorhergesagte Wahrscheinlichkeit für die Wahl der rechten Extremkategorie ansteigt, umso höher die Einstellungsstärke wird (vgl. Modell 4 in Unterabschnitt 9.1.4). Dieses Ergebnis ist inhaltlich sinnvoll, denn Befragte mit einer starken Einstellung zu einem Konstrukt sollen diese mit einer starken Zustimmung oder Ablehnung zu den Items Ausdruck verleihen.

Die Frage, ob Befragte mit einer höheren Einstellungsstärke weniger empfänglich sind für die Effekte des Antwortskalenlayouts als Befragte mit einer niedrigen und mittleren Einstellungsstärke, wird mittels der Hinzunahme der Interaktion von Einstellungsstärke

und Layout beantwortet (vgl. Modell 5 in Unterabschnitt 9.1.5). Die Ergebnisse unterstützen die Annahme, dass die Unterschiede der Effekte auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie zwischen den Layoutgruppen des NO-DK- und des ADD-DK-Layouts im Bereich der niedrigen und mittleren Einstellungsstärke sehr deutlich sind, dann kleiner werden und schließlich im Bereich der hohen Einstellungsstärke nicht mehr vorhanden sind. Die Betrachtung der Effekte auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie nach Einstellungsstärke für die Gruppen SEP-DK- und ADD-DK-Layout zeigt nahezu keine signifikanten Effektunterschiede. Lediglich für einen engen Bereich der mittleren Einstellungsstärke sind Unterschiede in der Wahl der rechten Extremkategorie zu sehen.

Diese Ergebnisse sind in sehr ähnlicher Form auch bei der nach Item-Konstrukt getrennten Betrachtung der Effekte auf die Wahrscheinlichkeiten zu finden. Wiederum zeigt sich, dass mit steigender Einstellungsstärke Befragte weniger empfänglich für Layouteffekte sind. Dieser Zusammenhang ist vor allem bei dem Vergleich des NO-DK- und ADD-DK-Layouts ersichtlich und wiederum bei den PES- stärker als bei den LIST-Items.

Tabelle 7 Ergebnisse der logistischen Multilevel-Modelle zur Wahl der rechten Extremkategorie

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	β	β	β	β	β
	(se, robust)	(se, robust)	(se, robust)	(se, robust)	(se, robust)
Layout					
NO-DK	-0,634*** (0,14)	-0,636*** (0,14)	-0,670*** (0,19)	-0,721*** (0,19)	-2,432*** (0,56)
SEP -DK	-0,260* (0,13)	-0,261* (0,13)	-,4667** (0,19)	-0,467* (0,19)	-0,504 (0,46)
Item-Konstrukt					
LIST		-1,034*** (0,31)	-1,184*** (0,33)	-1,189*** (0,33)	-1,185*** (0,33)
Konstrukt×Layout					
LIST×NO-DK			0,064 (0,27)	0,045 (0,27)	-0,009 (0,28)
LIST×SEP-DK			0,389 (0,26)	0,337 (0,26)	0,365 (0,26)
Einstellungsstärke					
				0,053*** (0,01)	0,022 (0,02)
ES⁺×Layout					
ES ⁺ ×NO-DK					0,102** (0,03)
ES ⁺ ×SEP-DK					0,002 (0,03)
Konstante	-1,981*** (0,15)	-1,563*** (0,18)	-1,487*** (0,19)	-2,295*** (0,27)	-1,817*** (0,36)
McFadden's-R ²	0,078	0,083	0,084	0,092	0,098
AIC	2534,49	2524,16	2525,65	2504,44	2491,36
BIC	2652,05	2648,25	2662,80	2648,12	2648,10
N	5070	5070	5070	5070	5070

Referenzkategorie: ADD-DK (Layout); PES (Item-Konstrukt); ⁺ES=Einstellungsstärke

* p<0,05, ** p<0,01, *** p<0,001

Tabelle 8 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie¹³

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)
Layout					
NO-DK	-0,042 (-0,06; -,02)	-0,042 (-0,06; -0,23)	-0,042 (-0,06; -0,02)	-0,045 (-0,06; -0,03)	-0,045 (-0,06; -0,03)
SEP-DK	-0,020 (-0,04; 0,00)	-0,020 (-0,04; 0,00)	-0,020 (-0,04; 0,00)	-0,022 (-0,04; 0,00)	-0,019 (-0,04; 0,00)
Item-Konstrukt					
LIST		-0,077 (-0,13; -0,02)	-0,077 (-0,13; -0,02)	-0,078 (-0,13; -0,03)	-0,079 (-0,13; -0,03)
Konstrukt×Layout					
PES×NO-DK			-0,073 (-0,12; -0,03)	-0,078 (-0,12; -0,03)	-0,080 (-0,12; -0,04)
PES×SEP-DK			-0,054 (-0,10; -0,01)	-0,055 (-0,10; -0,01)	-0,052 (-0,10; -0,01)
LIST×NO-DK			-0,027 (-0,05; -0,01)	-0,030 (-0,05; -0,01)	-0,030 (-0,05; -0,01)
LIST×SEP-DK			-0,004 (-0,02; 0,01)	-0,007 (-0,03; -0,01)	-0,004 (-0,02; 0,02)
Einstellungsstärke				0,003 (0,00; 0,01)	0,003 (0,00; 0,01)
N	5070	5070	5070	5070	5070

Referenzkategorie: ADD-DK (Layout); PES-Items (Item-Konstrukt), ⁺ KI = Konfidenzintervall

¹³ In Modell 5 sind die AMEs zu den 25 Stufen der Einstellungsstärke nach Skalenlayout aus Gründen der Übersichtlichkeit nicht einzeln aufgeführt. Sie können jedoch den Graphen in Abbildung 15 und in Abbildung 17 entnommen werden.

Tabelle 9 Vorhergesagte Wahrscheinlichkeiten (Whs.) der Wahl der rechten Extremkategorie¹⁴

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)
Layout					
NO-DK	0,055 (0,04; 0,07)	0,055 (0,04; 0,07)	0,055 (0,04; 0,07)	0,054 (0,04; 0,06)	0,055 (0,05; 0,07)
SEP-DK	0,077 (0,06; 0,09)	0,077 (0,06; 0,09)	0,077 (0,06; 0,09)	0,077 (0,07; 0,09)	0,076 (0,06; 0,09)
ADD-DK	0,097 (0,08; 0,11)	0,097 (0,08; 0,11)	0,097 (0,08; 0,11)	0,099 (0,09; 0,11)	0,097 (0,08; 0,11)
Item-Konstrukt					
PES-Items		0,129 (0,08; 0,17)	0,129 (0,08; 0,17)	0,130 (0,09; 0,17)	0,127 (0,08; 0,17)
LIST-Items		0,052 (0,04; 0,06)	0,052 (0,04; 0,06)	0,052 (0,04; 0,06)	0,052 (0,04; 0,06)

- Tabellenfortsetzung auf der nächsten Seite -

¹⁴ In Modell 4 und Modell 5 werden aus Gründen der Übersichtlichkeit nicht die vorhergesagten Wahrscheinlichkeiten der 25 Ausprägungen der Einstellungsstärke sowie der Werte der Einstellungsstärke nach Antwortskalenlayout aufgeführt. Diese Werte können aus den Graphen in Abbildung 14, Abbildung 16 und Abbildung 18 abgelesen werden.

- Fortsetzung der Tabelle -

	Modell 1 Whs. (95 % KI ⁺)	Modell 2 Whs. (95 % KI ⁺)	Modell 3 Whs. (95 % KI ⁺)	Modell 4 Whs. (95 % KI ⁺)	Modell 5 Whs. (95 % KI ⁺)
Konstrukt×Layout					
PES×NO-DK			0,100 (0,06; 0,14)	0,099 (0,06; 0,14)	0,097 (0,06; 0,14)
PES×SEP-DK			0,119 (0,07; 0,17)	0,122 (0,07; 0,17)	0,125 (0,08; 0,18)
PES×ADD-DK			0,173 (0,11; 0,23)	0,177 (0,12; 0,24)	0,177 (0,12; 0,24)
LIST×NO-DK			0,036 (0,02; 0,05)	0,035 (0,02; 0,05)	0,033 (0,02; 0,04)
LIST×SEP-DK			0,059 (0,04; 0,08)	0,058 (0,04; 0,07)	0,059 (0,04; 0,08)
LIST×ADD-DK			0,063 (0,05; 0,08)	0,065 (0,05; 0,08)	0,063 (0,05; 0,08)
N	5070	5070	5070	5070	5070

⁺ KI = Konfidenzintervall

9.2 Wahl der Mittelkategorie

Dieser Abschnitt befasst sich mit den Ergebnissen der fünf logistischen Multilevel-Modelle der Wahl der Mittelkategorie. Dabei gibt die zu erklärende Variable „Wahl der Mittelkategorie“ für jedes der siebzehn untersuchten Items der PES und des LIST-Inventars an, ob ein Befragter zu dessen Beantwortung die Mittelkategorie gewählt oder nicht gewählt hat. Somit wird die mittlere Neigung aller Befragten untersucht, in Abhängigkeit der in den Modellen betrachteten erklärenden Variablen die Mittelkategorie zur Fragebeantwortung zu nutzen.

9.2.1 Modell 1: Effekt des Layouts

Um den Effekt des Antwortskalenlayouts auf die Wahrscheinlichkeit der Wahl der Mittelkategorie zu untersuchen, wird in Modell 1 die nominale Variable „Skalenlayout“ als Prädiktor einbezogen.

Die Logit-Koeffizienten dieser nominalen Variablen sind sowohl für die Gruppe des NO-DK-Layouts als auch für die Gruppe des SEP-DK-Layouts, in Referenz zum ADD-DK-Layout, signifikant positiv. Demnach wählen Befragte in diesen beiden Layouts die Mittelkategorie wahrscheinlicher als solche im ADD-DK-Layout. Gemäß des McFaddens- R^2 von 0,027 kommt es durch die Berücksichtigung der Variable „Skalenlayout“ zu einer leichten Steigerung der Erklärungskraft im Vergleich zu einem Modell ohne Prädiktoren. Die Betrachtung der AMEs, dargestellt im Graphen in Abbildung 19, verdeutlicht ebenfalls, dass die Effekte auf die Wahrscheinlichkeit einer mittleren Antwortabgabe im NO-DK-Layout um 5 Prozentpunkte und im SEP-DK-Layout um 3 Prozentpunkte signifikant stärker sind als der Effekt des ADD-DK-Layouts (vgl. auch Tabelle 11, Seite 120).

Wird im gleichen Multilevel-Modell das NO-DK-Layout als Referenzkategorie definiert, zeigt der AME, dass sich das NO-DK- und SEP-DK-Layout nicht signifikant in Bezug auf die Wahl der Mittelkategorie voneinander unterscheiden.

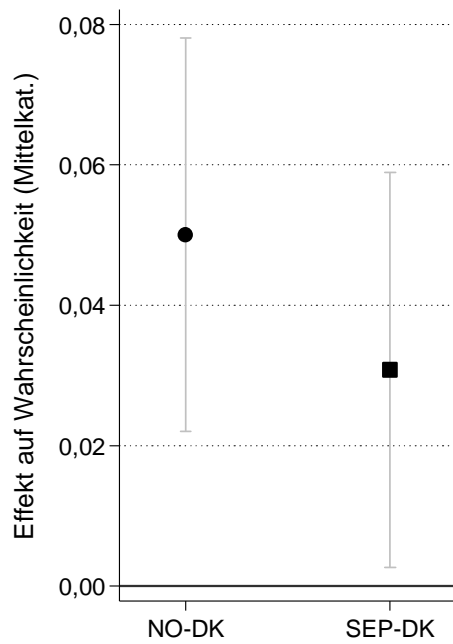


Abbildung 19 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK)

Diese Effektunterschiede finden Ausdruck in den vorhergesagten Wahrscheinlichkeiten der Wahl der Mittelkategorie (vgl. Abbildung 20):

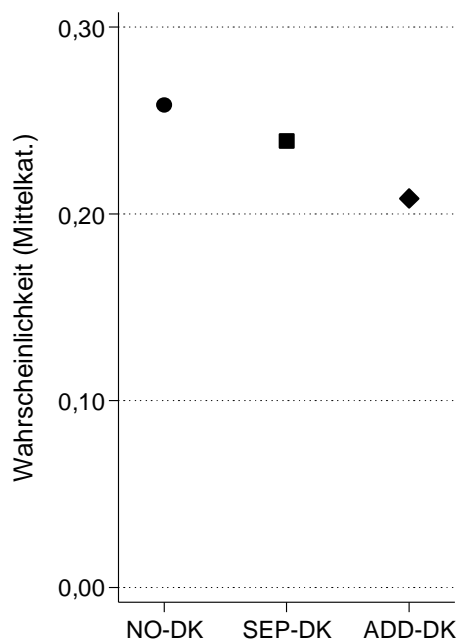


Abbildung 20 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout

Für das NO-DK-Layout wird mit 26 % eine höhere Wahrscheinlichkeit als mit 21 % für das ADD-DK-Layout vorhergesagt. Die Wahrscheinlichkeit der Wahl der Mittelkategorie für das SEP-DK-Layout liegt mit 24 % dazwischen.

9.2.2 Modell 2: Effekt des Item-Konstrukts

In Modell 2 wird zusätzlich die nominale Variable „Item-Konstrukt“ aufgenommen, die den untersuchten Items zuweist, welches der Konstrukte „PES“ bzw. „LIST“ sie repräsentieren.

Der Logit-Koeffizient für das LIST-Konstrukt nimmt einen signifikanten, positiven Wert an, was besagt, dass Befragte bei der Beantwortung der LIST-Items wahrscheinlicher die Mittelkategorie zur Antwortabgabe auswählen als bei der Beantwortung der PES-Items. Die Stärke des Unterschieds zwischen dem Effekt auf die Wahrscheinlichkeit der Wahl der Mittelkategorie des PES- und des LIST-Konstrukts beträgt gemäß der AMEs 17 Prozentpunkte und ist signifikant (vgl. Tabelle 11, Seite 120).

Dieser Unterschied spiegelt sich in den vorhergesagten Wahrscheinlichkeiten für die Wahl der Mittelkategorie wider: Befragte wählen mit 14,3 % bei der Beantwortung der PES-Items gerade mal halb so wahrscheinlich diese Kategorie zur Antwortabgabe wie mit 31,3 % bei der Beantwortung der LIST-Items (vgl. Tabelle 12, Seite 121).

Eine Erklärung dafür, dass Befragte bei den LIST-Items im Vergleich zu den PES-Items deutlich wahrscheinlicher die Mittelkategorie zur Fragebeantwortung auswählen, kann in der unterschiedlichen Dimensionalität der beiden Antwortskalen der beiden Item-Konstrukte begründet sein. Die Antwortskala der LIST-Items ist unipolar. Der Befragte soll darauf abstimmen, inwieweit die verschiedenen Lernsituationen auf ihn „überhaupt nicht“ bis „voll und ganz“ zutreffen. Die Mittelkategorie steht für ein „teilweises“ Zutreffen der beschriebenen Lernstrategie. Bei den PES-Items hingegen wird den Befragten eine bipolare Antwortskala vorgelegt. Auf dieser sollen sie eine Einstufung vornehmen, inwieweit sie die einzelnen Aussagen über die politische Einflussnahme „stark“ ablehnen bis ihnen „stark“ zustimmen. Auf dieser Antwortskala steht die Mittelkategorie für eine

indifferente Position („weder noch“). Es ist anzunehmen, dass die meisten Befragten für die einzelnen Items der PES eine Einstellung besitzen, die entweder eher in Richtung Ablehnung oder Zustimmung neigt und sie daher nicht eine indifferente Position über die Mittelkategorie zum Ausdruck bringen.

Im Modellvergleich werden keine Unterschiede hinsichtlich der Effekte der weiteren erklärenden Variablen zwischen Modell 1 und Modell 2 ersichtlich. Auch unter Kontrolle der Konstrukt-Zugehörigkeit der einzelnen Items ist die Wahrscheinlichkeit im NO-DK-Layout und im SEP-DK-Layout signifikant höher, die Mittelkategorie zur Antwortabgabe auszuwählen als im ADD-DK-Layout. Der leichte Anstieg von McFaddens- R^2 von 0,027 (Modell 1) auf 0,032 (Modell 2) sowie die Verringerungen der Werte des AIC und des BIC sprechen dafür, dass die Hinzunahme der erklärenden Variable „Item-Konstrukt“ die Erklärungskraft des Modells verbessert (vgl. Tabelle 10, Seite 119).

9.2.3 Modell 3: Interaktionseffekt „Item-Konstrukt \times Layout“

In Modell 3 wird der Interaktionsterm „Item-Konstrukt \times Layout“ in die Analyse einbezogen, um die Annahme der Hypothese (H 2), dass das Layout Befragte bei der Beantwortung beider Item-Konstrukte unterschiedlich in der Wahl der Mittelkategorie beeinflusst, zu testen.

Die beiden positiven Logit-Koeffizienten der Interaktion sind nicht signifikant. Die AMEs (vgl. Abbildung 21) zeigen jedoch, dass sich die Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie der Skalenlayouts zwischen den beiden Item-Konstrukten unterscheiden: Bei den PES-Items sind lediglich für das NO-DK-Layout in Bezug zum ADD-DK-Layout signifikant höhere Effekte ($AME = 3,6$) auf die Wahrscheinlichkeit zu beobachten. Bei den LIST-Items unterscheiden sich sowohl die Effekte des NO-DK-Layouts ($AME = 5,9$) als auch des SEP-DK-Layouts ($AME = 4,3$) signifikant von dem des ADD-DK-Layouts (vgl. Tabelle 11, Seite 120).

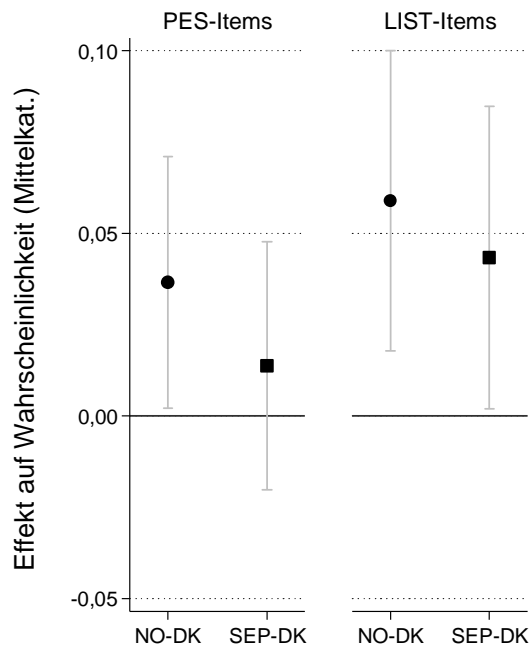


Abbildung 21 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Item-Konstrukt

Die nach Item-Konstrukt getrennte Betrachtung der vorhergesagten Wahrscheinlichkeiten (vgl. Abbildung 22) zeigt bei den PES-Items für alle Layoutgruppen mit 12,6 % (ADD-DK) bis 16,2 % (NO-DK) deutlich geringere vorhergesagte Wahrscheinlichkeiten für die Wahl der Mittelkategorie als für die LIST-Items mit 27,7 % (ADD-DK) bis 33,6 % (NO-DK) (vgl. Tabelle 12, Seite 121).¹⁵

¹⁵ Die unterschiedliche Dimensionalität der beiden Ratingskalen (unipolar vs. bipolar) als mögliche Ursache für diese Niveauunterschiede wurde im vorangehenden Unterabschnitt (9.2.2) diskutiert.

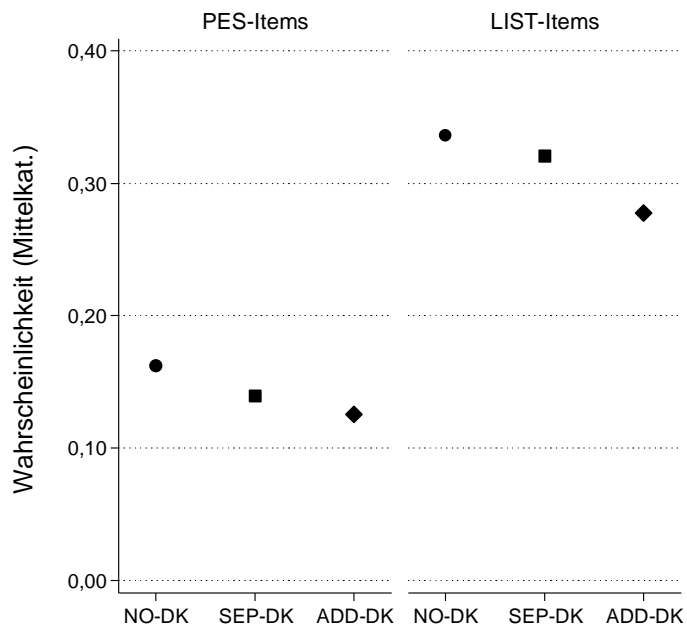


Abbildung 22 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout und Item-Konstrukt

Im Vergleich zum vorhergehenden Modell 2 verliert der Logit-Koeffizient für das SEP-DK-Layout in Modell 3 unter Hinzunahme des Interaktionsterms „Item-Konstrukt \times Layout“ seine Signifikanz. Die Effekte der übrigen Prädiktoren sind zwischen den Modellen unverändert. Demnach bleibt die höhere Wahrscheinlichkeit einer Antwortabgabe anhand der Mittelkategorie in der Gruppe des NO-DK-Layouts gegenüber dem ADD-DK-Layout bestehen. Genauso zeigt sich weiterhin eine niedrigere Wahrscheinlichkeit für die Wahl der Mittelkategorie bei Beantwortung der PES-Items im Vergleich zu den LIST-Items. Zwar geht aus den zuvor beschriebenen Ergebnissen hervor, dass für die LIST-Items stärkere Unterschiede in der Wahl der Mittelkategorie zwischen den drei Layoutgruppen beobachtet werden können als bei den PES, trotzdem trägt diese Berücksichtigung der Interaktion nicht zu einer Verbesserung der Erklärungskraft des Modells bei. Weder steigt McFaddens- R^2 an, noch verringern sich die Werte des AIC und des BIC (vgl. Tabelle 10, Seite 119).

9.2.4 Modell 4: Effekt der Einstellungsstärke

In Modell 4 wird die individuelle Einstellungsstärke gegenüber den Themen der beiden Item-Konstrukte „PES“ und „LIST“ in die Analyse miteinbezogen.

Die Ergebnistabelle der logistischen Multilevel-Regression weist einen signifikanten, negativen Logit-Koeffizienten für die Einstellungsstärke aus. Demnach nimmt über alle untersuchten Items hinweg die Wahrscheinlichkeit der Wahl der Mittelkategorie mit steigender Einstellungsstärke ab. Laut dem AME nimmt dieser negative Zusammenhang folgende Stärke an: Steigt die Einstellungsstärke um eine Einheit, so nimmt die Wahrscheinlichkeit der Wahl der Mittelkategorie um 0,4 Prozentpunkte ab (vgl. Tabelle 11, Seite 120). Es wird für Befragte mit einer sehr geringen Einstellungsstärke (1 Punkt) eine Wahrscheinlichkeit von rund 30 % vorhergesagt, die Mittelkategorie zu wählen (vgl. Abbildung 23). Diese verringert sich auf rund 20 % für Befragte, die angeben, gegenüber den thematischen Inhalten der Items eine hohe Einstellungsstärke (25 Punkte) zu besitzen. Dieser negative Zusammenhang ist inhaltlich sinnvoll, denn eine starke Einstellung sollte durch Antworten zu Kategorien links oder rechts der Skalenmitte zum Ausdruck gebracht werden.

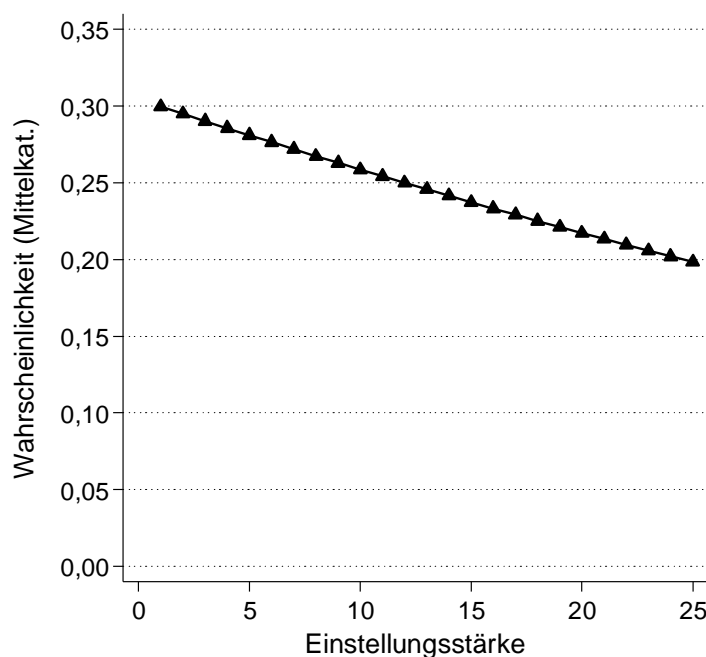


Abbildung 23 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Einstellungsstärke

Der Vergleich der Effekte der übrigen erklärenden Variablen zwischen Modell 4 und dem vorangehenden Modell 3 zeigt keine Veränderungen (vgl. Tabelle 11, Seite 120). Befragte im NO-DK-Layout haben eine signifikant höhere Wahrscheinlichkeit, die Mittelkategorie zu wählen als Befragte im ADD-DK-

Layout. Des Weiteren ist die Wahrscheinlichkeit einer Antwortabgabe anhand der Mittelkategorie für die LIST-items wahrscheinlicher als für die PES-Items. Der leichte Anstieg von McFaddens- R^2 von 0,032 (Modell 3) auf 0,034 (Modell 4) und die Verringerungen der Werte des AIC und des BIC sprechen zudem dafür, dass die Hinzunahme der Variable „Einstellungsstärke“ die Erklärungskraft dieses Modells gegenüber dem vorangehenden Modell 3 anhebt (vgl. Tabelle 10, Seite 119).

9.2.5 Modell 5: Interaktionseffekt „Einstellungsstärke × Layout“

Im letzten Modell 5 zur Wahl der Mittelkategorie wird der Interaktionsterm „Einstellungsstärke × Layout“ in die logistische Multilevel-Regression mitaufgenommen. Damit wird es möglich, die Hypothese (H 3) zu untersuchen, deren Annahme lautet, dass Befragte in der Wahl der Antwortkategorie mit steigender Einstellungsstärke weniger durch das Layout der Ratingskala beeinflusst werden.

Die beiden Logit-Koeffizienten für die Interaktionen der Einstellungsstärke mit dem NO-DK- bzw. SEP-DK-Layout (in Referenz zum ADD-DK-Layout) sind negativ und nicht signifikant (vgl. Tabelle 10, Seite 119). Aus der Betrachtung der graphischen Darstellung der AMEs (vgl. Abbildung 24) geht hervor, dass sowohl im NO-DK-Layout als auch im SEP-DK-Layout, in Bezug zum ADD-DK-Layout, mit steigender Einstellungsstärke die Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie kleiner werden. Die AMEs des NO-DK-Layouts unterscheiden sich im Bereich einer niedrigen bis mittleren Einstellungsstärke (1 bis 19 Punkte) signifikant von denen des ADD-DK-Layouts. Im Bereich der hohen Einstellungsstärke (ab 20 Punkten) gehen die AMEs des NO-DK-Layouts gegen Null und unterscheiden sich nicht mehr signifikant von denen des ADD-DK-Layouts. Zwischen dem SEP-DK- und dem ADD-DK-Layout werden signifikante Effektunterschiede nur für den Bereich einer mittleren Einstellungsstärke (10 bis 16 Punkte) beschrieben. Zwar sind für den Bereich der niedrigen Einstellungsstärke (1 bis 9 Punkte) stärkere Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie ausgewiesen als für den Bereich der mittleren Einstellungsstärke (10 bis 16 Punkte), allerdings werden diese aufgrund ihrer breiten Konfidenzintervalle als nicht signifikant unterschiedlich von den Effekten der

Referenzgruppe bewertet. Die breiten Konfidenzintervalle resultieren aus den geringen Fallzahlen im unteren Bereich der Einstellungsstärke, welche hohe Standardfehler verursachen. Demnach kann aus diesen Ergebnissen nur mit einer großen Unsicherheit geschlossen werden, dass die Effekte auf die Wahrscheinlichkeit im SEP-DK-Layout und ADD-DK-Layout sich nicht signifikant voneinander unterscheiden, was bei der Ergebnisinterpretation berücksichtigt werden sollte.

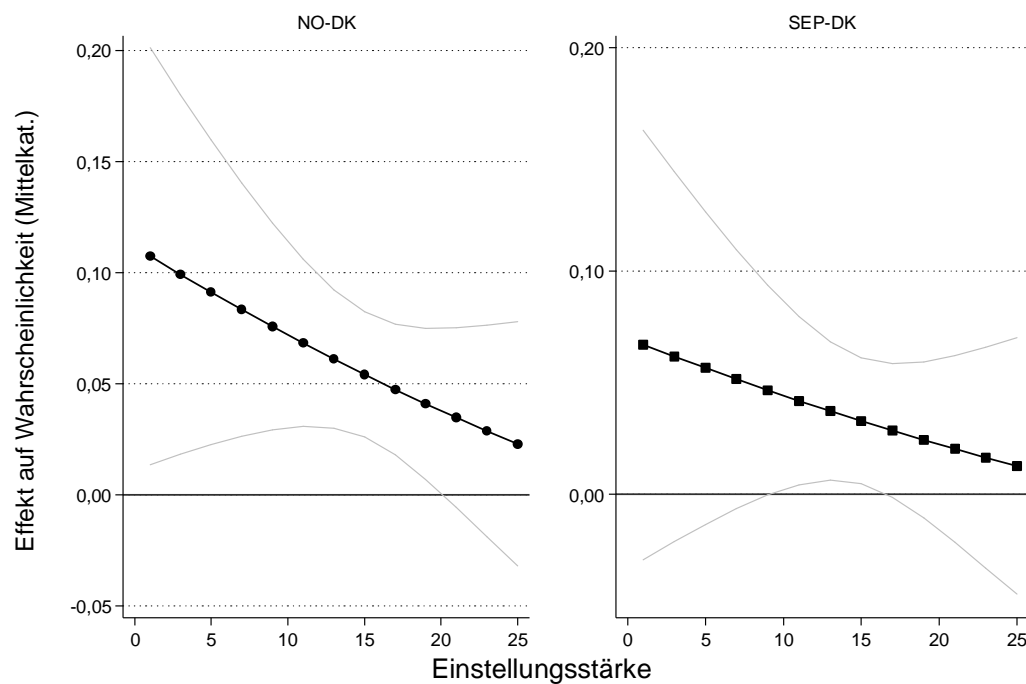


Abbildung 24 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Einstellungsstärke

Aus den vorhergesagten Wahrscheinlichkeiten für die drei Antwortskalenlayouts geht hervor, dass im Bereich einer niedrigen bis mittleren Einstellungsstärke (1 bis 19 Punkte) sowohl im NO-DK-Layout als auch im SEP-DK-Layout deutlich wahrscheinlicher die Mittelkategorie gewählt wird als im selben Einstellungsstärkebereich im ADD-DK-Layout (vgl. Abbildung 25). Die stärkste negative Steigung ist dabei im NO-DK-Layout zu beobachten. Hier fällt die vorhergesagte Wahrscheinlichkeit der Wahl der Mittelkategorie um 14 Prozentpunkte, nämlich von 34,7 % für Befragte mit einer sehr niedrigen Einstellungsstärke (1 Punkt) auf 20,9 % für Befragte mit einer sehr hohen Einstellungsstärke (25 Punkte). Im SEP-DK-Layout ist eine etwas geringere, aber immer noch deutliche Abnahme der Wahrscheinlichkeit um 10 Prozentpunkte mit

steigender Einstellungsstärke zu beobachten, nämlich von 30,7 % bei einer sehr niedrigen Einstellungsstärke (1 Punkt) auf 20 % bei einer sehr hohen Einstellungsstärke (25 Punkte). Ein anderes Bild zeigt sich hingegen für die Kurve des ADD-DK-Layouts: Obwohl auch hier ein negativer Zusammenhang zwischen der Wahrscheinlichkeit der Wahl der Mittelkategorie und der Einstellungsstärke zu sehen ist, beträgt die Abnahme von 24 % bei einer sehr niedrigen Einstellungsstärke (1 Punkt) auf 18,6 % bei einer sehr hohen Einstellungsstärke (25 Punkte) mit gut 5 Prozentpunkten Differenz lediglich etwa ein Drittel im Vergleich zur Abnahme der Wahrscheinlichkeit im NO-DK-Layout und die Hälfte im Vergleich zum SEP-DK-Layout. Im Bereich der hohen Einstellungsstärke unterscheiden sich die vorhergesagten Wahrscheinlichkeiten der drei Layoutgruppen nicht mehr maßgeblich voneinander, sie liegt jeweils bei circa 20 %.

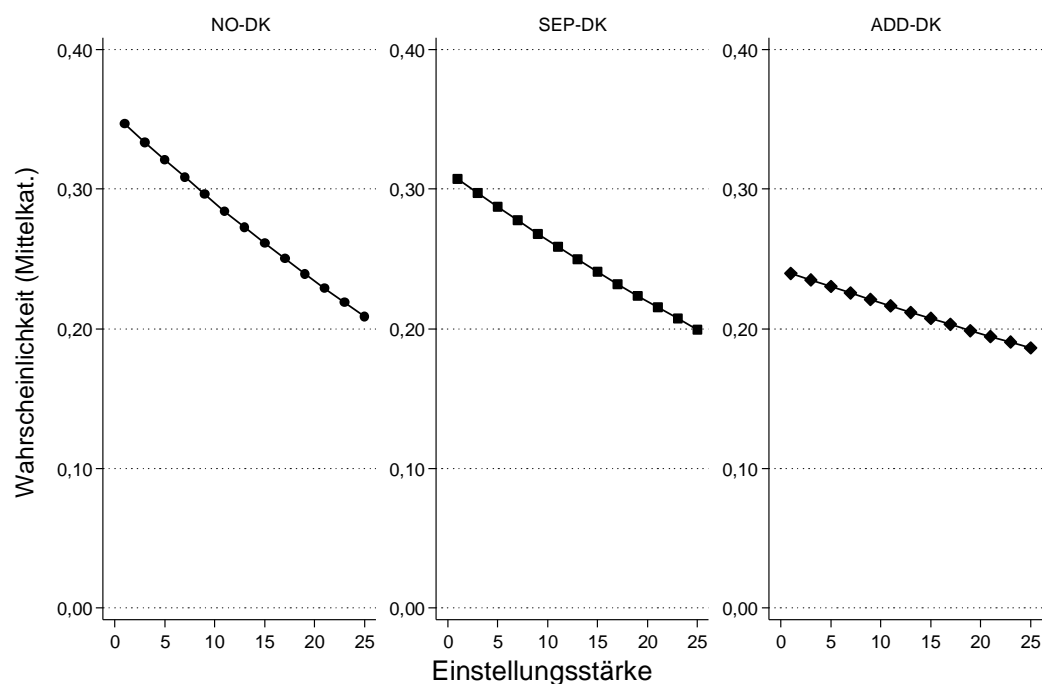


Abbildung 25 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout und Einstellungsstärke

Da aus dem vorherigen Modell 4 hervorgeht, dass Befragte bei der Beantwortung der PES-Items weniger wahrscheinlich die Mittelkategorie wählen als bei der Beantwortung der LIST-Items, wird hier auch die Interaktion von Einstellungsstärke und Skalenlayout getrennt für die beiden Item-Konstrukte betrachtet (vgl. Abbildung 26). Hierbei ist ein ähnliches Bild wie bei der gemeinsamen Analyse der Interaktion „Einstellungsstärke \times Layout“ über alle Items

zu erkennen: Die AMEs, sowohl für die Items der PES als auch des LIST-Konstrukts, zeigen signifikante Unterschiede der Effekte auf die Wahrscheinlichkeit zwischen dem NO-DK-Layout und dem ADD-DK-Layout im unteren bis mittleren Bereich der Einstellungsstärke, aber nicht mehr für den höheren Einstellungsstärkebereich. Insgesamt sind die Effektunterschiede zwischen den Layouts bei den LIST-Items 3 bis 4 Prozentpunkte höher als bei den PES-Items. Bei dem Vergleich des SEP-DK- und ADD-DK-Layouts werden nur für die LIST-Items und nur im Bereich der mittleren Einstellungsstärke (8 bis 16 Punkte) signifikant unterschiedliche Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie nach Einstellungsstärke gefunden.

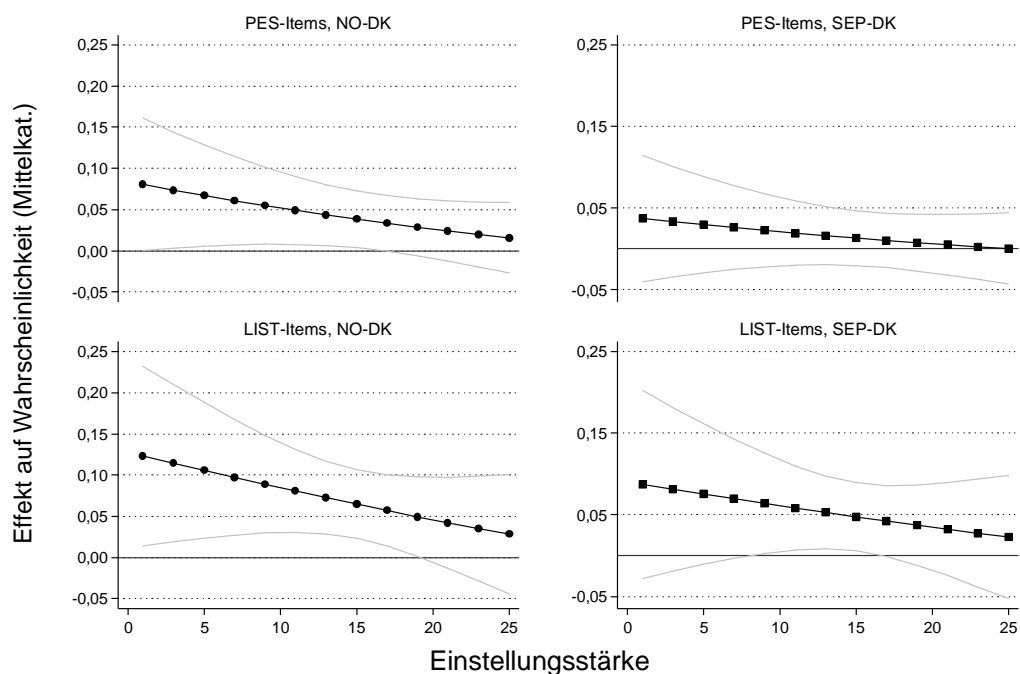


Abbildung 26 Marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK), Einstellungsstärke und Item-Konstrukt

Allerdings sei angemerkt, dass hier wieder das Problem der breiten Konfidenzintervalle der AMEs für den Bereich der niedrigen Einstellungsstärke besteht. Für die PES-Items unterscheiden sich die AMEs des SEP-DK-Layouts über alle Stufen der Einstellungsstärke nicht signifikant von dem Effekt der Referenzkategorie „ADD-DK-Layout“.

Schließlich veranschaulicht noch die Betrachtung der graphischen Darstellung der vorhergesagten Wahrscheinlichkeiten (vgl. Abbildung 27), dass auch nach der

Einbeziehung der Interaktion „Einstellungsstärke \times Layout“, wie zuvor bereits in Modell 4 beschrieben, bei der Beantwortung der PES-Items die Wahrscheinlichkeit für eine Antwortabgabe anhand der Mittelkategorie deutlich geringer ist als bei den LIST-Items. Für beide Item-Konstrukte sind zwischen den drei Layoutgruppen Unterschiede in der Wahrscheinlichkeit der Wahl der Mittelkategorie im unteren Bereich der Einstellungsstärke zu sehen, die mit zunehmender Einstellungsstärke geringer werden, bis sie im Bereich der sehr hohen Einstellungsstärke nahezu verschwunden sind. Die Kurven der vorhergesagten Wahrscheinlichkeiten nach Einstellungsstärke haben für die LIST-Items für das NO-DK- und SEP-DK-Layout steilere negative Steigungen als für die PES-Items. In beiden Item-Konstrukten sind die Steigungen der Gruppe des ADD-DK-Layouts im Vergleich zu den beiden anderen Layouts flacher. Bei den LIST-Items beträgt der Unterschied in der Wahrscheinlichkeit einer Wahl der Mittelkategorie im NO-DK-Layout 16 Prozentpunkte zwischen der niedrigsten und der höchsten Ausprägung der Einstellungsstärke, bei den PES-Items lediglich 10 Prozentpunkte. Im SEP-DK-Layout beträgt der Unterschied bei den LIST-Items 13 Prozentpunkte und bei den PES-Items 7 Prozentpunkte und im ADD-DK-Layout 7 Prozentpunkte bei den LIST- und 4 Prozentpunkte bei den PES-Items.

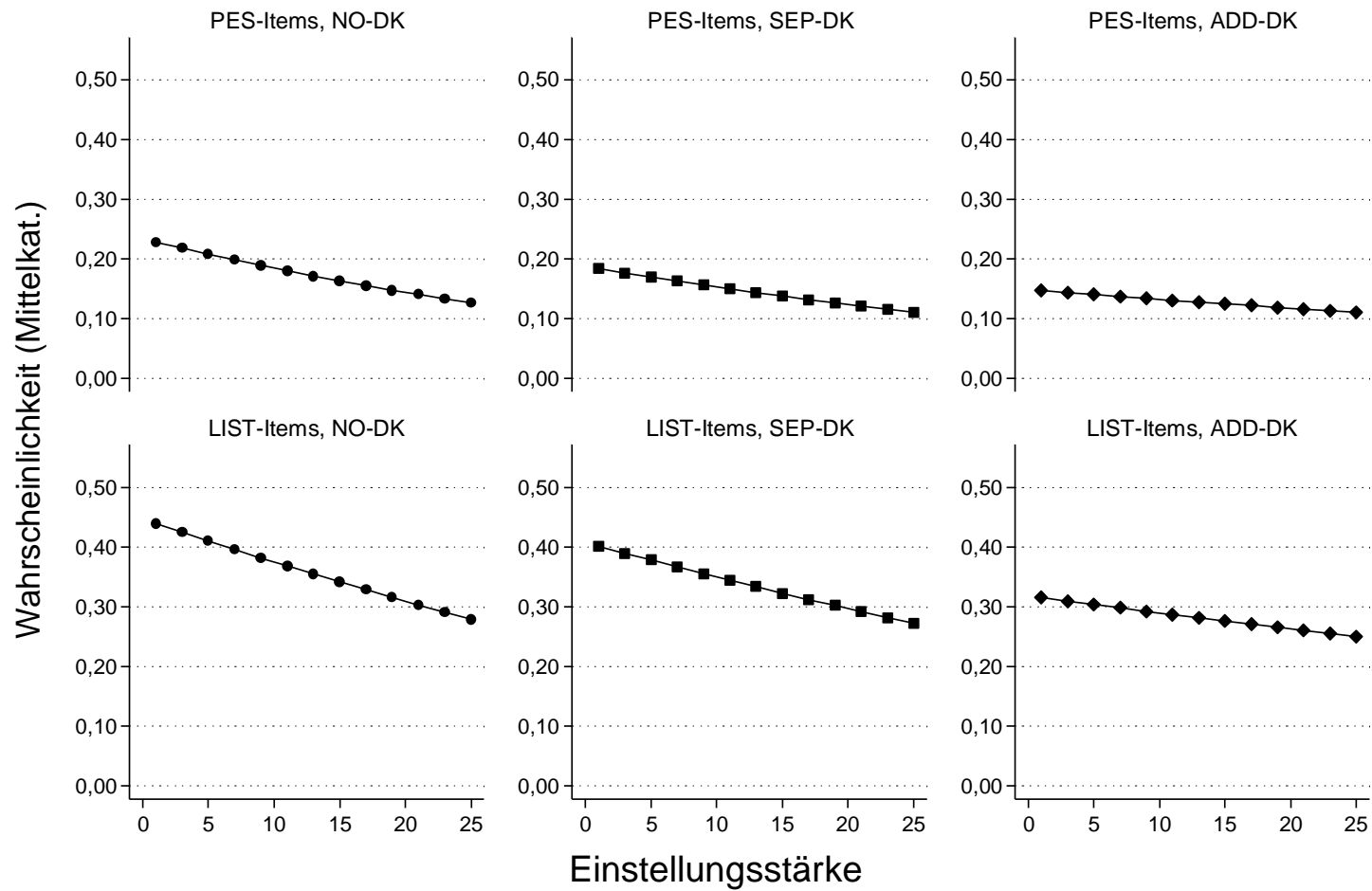


Abbildung 27 Vorhergesagte Wahrscheinlichkeiten der Wahl der Mittelkategorie nach Layout, Einstellungsstärke und Item-Konstrukt

9.2.6 Zwischenfazit

Effekt des Layouts der Ratingskala

Die Hypothese (H 1.2) über den Einfluss des Skalenlayouts lautet, dass Befragte im ADD-DK-Layout eine niedrigere vorhergesagte Wahrscheinlichkeit besitzen, die Mittelkategorie zu wählen als Befragte im NO-DK-Layout und SEP-DK-Layout. Begründet wird diese Annahme durch die klare Erkennbarkeit der Mittelkategorie im NO-DK-Layout und im SEP-DK-Layout, da hier eine Übereinstimmung des visuellen und des inhaltlichen Skalenmittelpunkts vorliegt. Dadurch können Befragte zu dieser Kategorie hingezogen werden, sei es als inhaltliche Antwort oder aus Gründen eines Satisficing-Verhaltens (vgl. Unterabschnitt 5.3).

Die Ergebnisse unterstützen diese Hypothese: Das Skalenlayout beeinflusst die Wahrscheinlichkeit der Wahl der Mittelkategorie des Befragten signifikant (vgl. Modell 1 in Unterabschnitt 9.2.1). Wie angenommen sind die Effekte auf die Wahrscheinlichkeit im NO-DK-Layout und im SEP-DK-Layout signifikant höher als im ADD-DK-Layout. Die höchsten Wahrscheinlichkeiten werden für das NO-DK-Layout, gefolgt vom SEP-DK-Layout und ADD-DK-Layout vorhergesagt. Die Kombination einer klar erkennbaren Mittelkategorie mit einer nicht vorgegebenen DK-Kategorie scheint Befragte im NO-DK-Layout besonders wahrscheinlich die Mittelkategorie wählen zu lassen.

Effekt des Item-Konstrukts

Die nach Item-Konstrukt getrennte Betrachtung veranschaulicht, dass Befragte bei der Beantwortung der LIST-Items doppelt so wahrscheinlich die Mittelkategorie wählen wie bei der Beantwortung der PES-Items (vgl. Modell 2 in Unterabschnitt 9.2.2). Dieser Unterschied zwischen den beiden Item-Konstrukten kann zum einen inhaltlich erklärt werden: Befragte wählen zur Beantwortung der LIST-Items wahrscheinlicher die Mittelkategorie „trifft teilweise zu“, da sie damit eine mittlere Einstellung gegenüber dem Konstrukt ausdrücken. Solch eine mittlere Einstellung ist bei vielen Befragten zu erwarten, was sich in der multivariaten Analyse in einer eher höheren vorhergesagten Wahrscheinlichkeit widerspiegelt. Bei der

Beantwortung der PES-Items hingegen beschreibt die Mittelkategorie „weder noch“ eine indifferente Position gegenüber den PES-Items. Es wird jedoch von den meisten Befragten erwartet, dass sie in ihrer Einstellung eher in eine ablehnende oder zustimmende Richtung neigen und somit nicht die Indifferenz ausdrückende Mittelkategorie wählen. Dies spiegelt sich in der multivariaten Analyse in niedrigeren vorhergesagten Wahrscheinlichkeiten wider.

Des Weiteren stellt sich auch die Frage, inwieweit die höhere Wahrscheinlichkeit der Wahl der Mittelkategorie bei den LIST-Items durch einen Konstrukt spezifischen Layouteffekt hervorgerufen wird (Hypothese H 2). Um diese Frage zu untersuchen, wird eine Interaktion in die logistische Multilevel-Regression einbezogen (vgl. Modell 3 in Unterabschnitt 9.2.3). Es zeigt sich, dass Befragte bei der Beantwortung der LIST-Items signifikant durch das Skalenlayout in der Wahl der Mittelkategorie beeinflusst werden: Sowohl im NO-DK-Layout als auch im SEP-DK-Layout wählen sie diese Kategorie zur Antwortabgabe wahrscheinlicher aus als im ADD-DK-Layout. Für die PES-Items wird lediglich ein signifikanter Layouteffekt für den Vergleich zwischen dem NO-DK-Layout und dem ADD-DK-Layout beobachtet. Diese Ergebnisse sprechen gegen die Annahme der Hypothese (H 2), die besagt, dass Befragte bei dem kognitiven Prozess der Beantwortung der abstrakteren PES-Items empfänglicher für die Layouteffekte sind als bei Beantwortung der LIST Items (vgl. Unterabschnitt 5.3).

Effekt der Einstellungsstärke

Für den Einfluss der Einstellungsstärke der Befragten zu den abgefragten Themen auf ihre Empfänglichkeit für das Skalenlayout wird in Hypothese (H 3) die Annahme formuliert, dass mit steigender Einstellungsstärke Befragte weniger durch den Fragekontext – und somit durch das Layout der Ratingskala - beeinflusst werden. Daher sollten Layouteffekte nur für Befragte mit einer niedrigen bis mittleren Einstellungsstärke zu beobachten sein (vgl. Unterabschnitt 2.2.1 und Unterabschnitt 5.3).

Bevor diese Interaktion untersucht wird, wird zunächst in Modell 4 der Haupteffekt der Variable „Einstellungsstärke“ gegenüber den Themen der politischen

Selbstwirksamkeit und gegenüber den persönlichen Lernstrategien betrachtet (vgl. Unterabschnitt 9.2.4). Wie erwartet nimmt die Wahrscheinlichkeit einer Antwortabgabe anhand der Mittelkategorie mit steigender Einstellungsstärke ab. Dies ist inhaltlich sinnvoll, denn je stärker die Einstellung gegenüber dem abgefragten Thema ist, desto weniger sollten Befragte die Mittelkategorie zu Gunsten einer extremeren Kategorie links oder rechts von ihr wählen. Dieser negative Zusammenhang sollte unabhängig davon zu beobachten sein, ob die Mittelkategorie eine indifferente Position auf einer bipolaren oder eine mittlere Position auf einer unipolaren Skala repräsentiert.

Schließlich wird untersucht, ob die Einstellungsstärke moderierend auf das Auftreten von Layouteffekten wirkt, ob also eine Interaktion zwischen diesen Variablen vorliegt (vgl. Modell 5 in Unterabschnitt 9.2.5). Wie angenommen zeigt sich zwischen den Layoutgruppen NO-DK und ADD-DK im Bereich der niedrigen bis mittleren Einstellungsstärke ein sehr deutlicher, signifikanter Unterschied der Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie. Wie angenommen wird dieser Unterschied mit zunehmender Einstellungsstärke geringer und ist schließlich im Bereich einer hohen Einstellungsstärke nicht mehr vorhanden. Die Betrachtung der Effekte auf die Wahrscheinlichkeiten für die Wahl der Mittelkategorie nach Einstellungsstärke für die Layoutgruppen SEP-DK und ADD-DK zeigt hingegen keinen Layouteffekt. Lediglich in einem engen Bereich der mittleren Einstellungsstärke sind die Unterschiede in der Wahl der Mittelkategorie signifikant. Aufgrund der breiten Konfidenzintervalle kann aus diesem Ergebnis allerdings nicht sicher geschlossen werden, dass sich die Effekte im niedrigen Bereich der Einstellungsstärke nicht doch voneinander unterscheiden.

Die nach Item-Konstrukten getrennte Betrachtung des Einflusses der Einstellungsstärke auf das Auftreten von Layouteffekten kommt schließlich zu dem Ergebnis, dass bei den LIST-Items mit steigender Einstellungsstärke weniger Effekte der graphischen Gestaltung der Antwortskala beobachtet werden. Eine Verringerung der Unterschiede der Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie mit zunehmender Einstellungsstärke ist vor allem zwischen dem NO-DK- und dem ADD-DK Layout deutlich erkennbar. Bei Beantwortung der PES-Items zeigt sich ebenfalls, dass Befragte mit einer niedrigen bis mittleren

Einstellungsstärke im Vergleich von NO-DK- und ADD-DK-Layout in der Wahl der Mittelkategorie durch das Layout beeinflusst werden. Ein signifikanter Unterschied des Effekts des SEP-DK-Layouts gegenüber dem des ADD-DK-Layouts besteht hingegen für keinen Bereich der Einstellungsstärke.

Die Ergebnisse für das LIST-Konstrukt unterstützen die Hypothese (H 3), die Ergebnisse für die PES-Items zumindest teilweise. Insgesamt scheinen die Unterschiede in der Wahl der Mittelkategorie nicht nur durch die graphische Gestaltung der Ratingskala, sondern auch durch die Vorgabe bzw. Nicht-Vorgabe einer DK-Kategorie hervorgerufen zu werden. Bei den PES-Items ist kein unterschiedlicher Effekt auf die Wahrscheinlichkeit der Wahl der Mittelkategorie über alle Ausprägungen der Einstellungsstärke zwischen dem SEP-DK- und dem ADD-DK-Layout zu beobachten, obwohl diese beiden Layouts sich graphisch deutlich mit Hinblick auf die Erkennbarkeit der Mittelkategorie voneinander unterscheiden.

Tabelle 10 Ergebnisse der logistischen Multilevel-Modelle zur Wahl der Mittelkategorie

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	β	β	β	β	β
	(se, robust)	(se, robust)	(se, robust)	(se, robust)	(se, robust)
Layout					
NO-DK	0,288*** (0,08)	0,290*** (0,08)	0,301* (0,15)	0,320* (0,15)	0,560 (0,29)
SEP -DK	0,182* (0,09)	0,183* (0,09)	0,120 (0,15)	0,124 (0,15)	0,286 (0,30)
Item-Konstrukt					
LIST		1,022*** (0,21)	0,998*** (0,24)	0,999*** (0,24)	0,999*** (0,24)
Konstrukt×Layout					
LIST×NO-DK			-0,018 (0,18)	-0,007 (0,18)	-0,003 (0,18)
LIST×SEP-DK			0,091 (0,18)	0,113 (0,19)	0,115 (0,18)
Einstellungsstärke				-0,024*** (0,01)	-0,014 (0,01)
ES⁺×Layout					
ES ⁺ ×NO-DK					-0,016 (0,02)
ES ⁺ ×SEP-DK					-0,011 (0,02)
Konstante	-1,405*** (0,11)	-1,993*** (0,18)	-1,977*** (0,20)	-1,631*** (0,22)	-1,776*** (0,28)
McFadden's-R ²	0,027	0,032	0,032	0,034	0,034
AIC	5430,01	5406,68	5410,23	5399,83	5402,83
BIC	5547,57	5530,77	5547,39	5543,52	5559,58
N	5070	5070	5070	5070	5070

Referenzkategorie: ADD-DK (Layout); PES (Item-Konstrukt); ⁺ES=Einstellungsstärke

* p<0,05, ** p<0,01, *** p<0,001

Tabelle 11 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der Mittelkategorie¹⁶

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)
Layout					
NO-DK	0,050 (0,02; 0,08)	0,050 (0,02; 0,08)	0,050 (0,02; 0,08)	0,054 (0,03; 0,08)	0,054 (0,03; 0,08)
SEP-DK	0,031 (0,00; 0,06)	0,031 (0,00; 0,06)	0,031 (0,00; 0,06)	0,034 (0,01; 0,06)	0,033 (0,00; 0,06)
Item-Konstrukt					
LIST		0,169 (0,11; 0,23)	0,169 (0,11; 0,23)	0,171 (0,11; 0,23)	0,171 (0,11; 0,23)
Konstrukt×Layout					
PES×NO-DK			0,036 (0,00; 0,07)	0,039 (0,00; 0,07)	0,038 (0,00; 0,07)
PES×SEP-DK			0,014 (-0,02; 0,05)	0,014 (-0,02; 0,05)	0,013 (-0,02; 0,05)
LIST×NO-DK			0,059 (0,02; 0,10)	0,065 (0,02; 0,07)	0,064 (0,02; 0,11)
LIST×SEP-DK			0,043 (0,00; 0,08)	0,048 (0,01; 0,09)	0,047 (0,01; 0,09)
Einstellungsstärke				-0,004 (-0,01; 0,00)	-0,004 (-0,01; 0,00)
N	5070	5070	5070	5070	5070

Referenzkategorie: ADD-DK (Layout); PES-Items (Item-Konstrukt), ⁺ KI = Konfidenzintervall

¹⁶ In Modell 5 sind die AMEs zu den 25 Stufen der Einstellungsstärke nach Skalenlayout aus Gründen der Übersichtlichkeit nicht einzeln aufgeführt. Sie können jedoch den Graphen in Abbildung 24 und Abbildung 26 entnommen werden.

Tabelle 12 Vorhergesagte Wahrscheinlichkeiten (Whs.) der Wahl der Mittelkategorie¹⁷

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)
Layout					
NO-DK	0,258 (0,24; 0,28)	0,258 (0,24; 0,28)	0,258 (0,24; 0,28)	0,260 (0,24; 0,28)	0,260 (0,24; 0,28)
SEP-DK	0,239 (0,22; 0,26)	0,239 (0,22; 0,26)	0,239 (0,22; 0,26)	0,239 (0,22; 0,26)	0,239 (0,22; 0,26)
ADD-DK	0,208 (0,19; 0,23)	0,208 (0,19; 0,23)	0,208 (0,19; 0,23)	0,206 (0,19; 0,23)	0,206 (0,19; 0,23)
Item-Konstrukt					
PES-Items		0,143 (0,11; 0,17)	0,143 (0,11; 0,17)	0,143 (0,11; 0,17)	0,143 (0,11; 0,17)
LIST-Items		0,313 (0,27; 0,35)	0,313 (0,27; 0,35)	0,314 (0,27; 0,35)	0,314 (0,27; 0,35)

- Tabellenfortsetzung auf der nächsten Seite -

¹⁷ In Modell 4 und Modell 5 werden aus Gründen der Übersichtlichkeit nicht die vorhergesagten Wahrscheinlichkeiten der 25 Ausprägungen der Einstellungsstärke sowie der Werte der Einstellungsstärke nach Antwortskalenlayout aufgeführt. Diese Werte können aus den Graphen in Abbildung 23, Abbildung 25 und Abbildung 27Abbildung 23 abgelesen werden.

- Fortsetzung der Tabelle -

	Modell 1 Whs. (95 % KI ⁺)	Modell 2 Whs. (95 % KI ⁺)	Modell 3 Whs. (95 % KI ⁺)	Modell 4 Whs. (95 % KI ⁺)	Modell 5 Whs. (95 % KI ⁺)
Konstrukt×Layout					
PES×NO-DK			0,162 (0,13; 0,20)	0,163 (0,13; 0,20)	0,163 (0,13; 0,20)
PES×SEP-DK			0,139 (0,11; 0,17)	0,138 (0,10; 0,17)	0,138 (0,10; 0,17)
PES×ADD-DK			0,126 (0,09; 0,16)	0,124 (0,09; 0,16)	0,125 (0,09; 0,16)
LIST×NO-DK			0,336 (0,29; 0,38)	0,339 (0,29; 0,39)	0,340 (0,29; 0,39)
LIST×SEP-DK			0,320 (0,27; 0,37)	0,323 (0,28; 0,37)	0,323 (0,28; 0,37)
LIST×ADD-DK			0,277 (0,23; 0,32)	0,275 (0,23; 0,32)	0,276 (0,23; 0,32)
N	5070	5070	5070	5070	5070

⁺ KI = Konfidenzintervall

9.3 Wahl der DK-Kategorie

Dieser letzte Abschnitt behandelt die Wahl der DK-Kategorie. Für die folgenden fünf logistischen Multilevel-Modelle gibt die abhängige Variable „Wahl der DK-Kategorie“ - analog zu den zuvor behandelten abhängigen Variablen „Wahl der rechten Extremkategorie“ und „Wahl der Mittelkategorie“ - für jedes der siebzehn untersuchten Items der PES und des LIST-Inventars an, ob ein Befragter zu dessen Beantwortung die DK-Kategorie zur Antwortabgabe gewählt oder nicht gewählt hat. Somit wird die mittlere Neigung aller Befragter untersucht, die DK-Kategorie zur Fragebeantwortung zu nutzen, in Abhängigkeit der in den Modellen betrachteten erklärenden Variablen.¹⁸

9.3.1 Modell 1: Effekt des Layouts

In Modell 1 wird mittels einer nominalen Variable der Einfluss des Skalenlayouts auf die Wahrscheinlichkeit des Auftretens einer DK-Antwort über alle 17 experimentell variierten Items untersucht.

Die logistische Multilevel-Regression weist einen positiven Logit-Koeffizienten aus, welcher lediglich auf dem 10 %-Niveau signifikant ist (vgl. Tabelle 13, Seite 135). Der AME zeigt keine unterschiedlichen Effekte auf die Wahrscheinlichkeit der Wahl der DK-Kategorie für das SEP-DK-Layout und das ADD-DK-Layout (vgl. Tabelle 14, Seite 136). Nach McFaddens- R^2 von 0,072 kommt es durch die Berücksichtigung des Skalenlayouts zu einer Steigerung der Erklärungskraft im Vergleich zu einem Modell ohne Prädiktoren (vgl. Tabelle 13, Seite 135). Schließlich liegt auch die vorhergesagte Wahrscheinlichkeit für die Wahl der DK-Kategorie im SEP-DK-Layout mit 2,5 % nur geringfügig höher als im ADD-DK-Layout mit 1,6 % (vgl. Abbildung 28 und Tabelle 15, Seite 137).

¹⁸ Befragte des NO-DK-Layouts sind von der Analyse ausgeschlossen, da sie keine Werte für die abhängige Variable besitzen können.

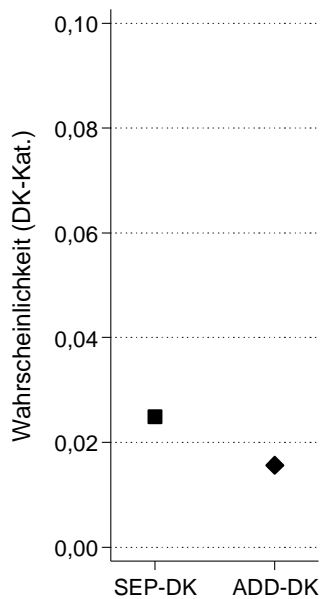


Abbildung 28 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout

9.3.2 Modell 2: Effekt des Item-Konstrukts

In Modell 2 wird unter Hinzunahme einer weiteren Indikatorvariable berücksichtigt, welches Item-Konstrukt von den einzelnen Items jeweils repräsentiert wird. Der Logit-Koeffizient für das LIST-Konstrukt nimmt einen negativen, nicht signifikanten Wert an (vgl. Tabelle 13, Seite 135). Auch die Betrachtung der nicht signifikanten AMEs zeigt, dass für die LIST-Items und die PES-Items keine unterschiedlichen Effekte auf die Wahrscheinlichkeit der Wahl der DK-Kategorie vorliegen (vgl. Tabelle 14, Seite 136). Die vorhergesagten Wahrscheinlichkeiten einer Wahl der DK-Kategorie sind auch bei einer nach Item-Konstrukt getrennten Betrachtung gering und betragen bei den PES-Items 2,7 % und bei den LIST-Items 1,3 % (vgl. Tabelle 15, Seite 137).

Auch unter Berücksichtigung der nominalen Variable „Item-Konstrukt“ in Modell 2 bleiben dessen weiteren Ergebnisse im Vergleich zum vorangehenden Modell 1 unverändert. Die Wahrscheinlichkeit der Wahl der DK-Kategorie unterscheidet sich nicht signifikant zwischen den beiden Layoutgruppen. Der leichte Anstieg des McFaddens- R^2 von 0,072 auf 0,082 weist darauf hin, dass die Hinzunahme der Indikatorvariable „Item-Konstrukt“ die Anpassung des Modells and die Daten leicht verbessert.

9.3.3 Modell 3: Interaktionseffekt „Item-Konstrukt \times Layout“

In Modell 3 ist von Interesse, ob Befragte bei der Beantwortung der Items zur Politischen Selbstwirksamkeit empfänglicher für Layouteffekte sind als bei der Beantwortung der Items zu ihren persönlichen Lernstrategien, wie in Hypothese (H 2) angenommen wird. Daher wird die Interaktion „Item-Konstrukt \times Layout“ in die logistische Multilevel-Regression aufgenommen. Der negative Logit-Koeffizient der Interaktion ist nicht signifikant. Allerdings wird nun unter Kontrolle dieser Interaktion der positive Logit-Koeffizient der nominalen Variable „Skalenslayout“ signifikant (vgl. Tabelle 13, Seite 135). Demnach wählen Befragte im SEP-DK-Layout, wie in Hypothese (H 1.3) angenommen, wahrscheinlicher die DK-Kategorie als Befragte im ADD-DK-Layout, wenn dafür kontrolliert wird, welches Item welchem Konstrukt zugehörig ist.

Bei der nach Item-Konstrukt getrennten Betrachtung der AMEs (vgl. Abbildung 29) wird erkennbar, dass bei Beantwortung der PES-Items zwischen den beiden Antwortskalenlayouts ein signifikanter Unterschied des Effekts auf die Wahrscheinlichkeit der Wahl der DK-Kategorie von 2 Prozentpunkten besteht. Ein signifikanter Layouteffekt auf die Wahl der DK-Kategorie bei Beantwortung der LIST-Items wird hingegen nicht beobachtet, hier kommt es zu einer Überschneidung des Konfidenzintervalls und der Nulllinie.

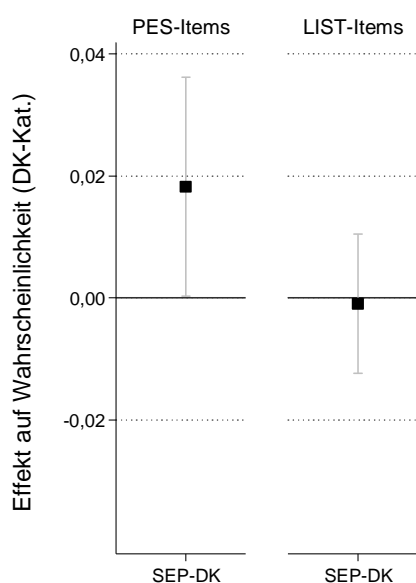


Abbildung 29 Durchschnittliche marginale Effekte (AMEs) der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Item-Konstrukt

Die vorhergesagte Wahrscheinlichkeit für die Wahl der DK-Kategorie (vgl. Abbildung 30 und Tabelle 15, Seite 137) beträgt für die PES-Items im SEP-DK-Layout 3,5 % und ist, wie erwartet, im ADD-DK-Layout mit 1,7 % geringer. Für die LIST-Items liegen für beide Layouts die vorhergesagten Wahrscheinlichkeiten bei 1,4 %. Nach diesem Ergebnis scheinen Befragte lediglich bei der Beantwortung der abstrakteren PES-Items durch das Layout in der Wahl der DK-Kategorie beeinflusst zu werden, bei der Beantwortung der LIST-Items hingegen nicht. Da es sich um einen Konstrukt spezifischen Layouteffekt handelt, wurde bei einer gemeinsamen Betrachtung in Modell 1 und Modell 2 der Layouteffekt nicht sichtbar.

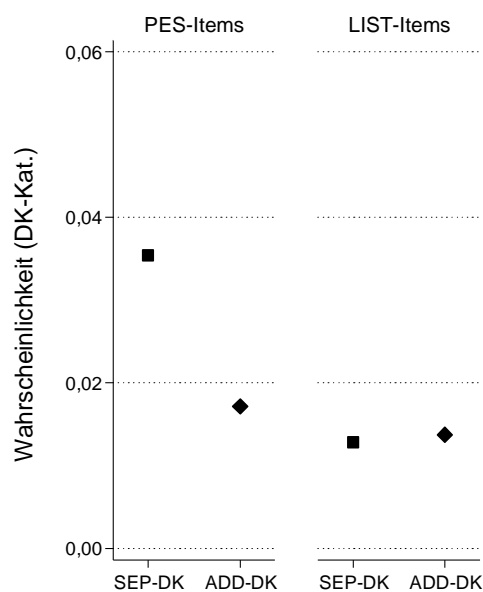


Abbildung 30 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout und Item-Konstrukt

Zwar führt die Berücksichtigung der Interaktion „Item-Konstrukt \times Layout“ dazu, dass ein signifikanter, Konstrukt spezifischer Layouteffekt für die PES-Items sichtbar wird, allerdings trägt dies nicht zu einer Verbesserung der Erklärungskraft des Modells bei. McFaddens- R^2 steigt geringfügig von 0,082 (Modell 2) auf 0,085 (Modell 3), die Werte des AIC und des BIC verringern sich nicht (vgl. Tabelle 13, Seite 135).

9.3.4 Modell 4: Effekt der Einstellungsstärke

Der Einfluss der individuellen Einstellungsstärke gegenüber den beiden abgefragten Themen „Politische Selbstwirksamkeit“ und „persönliche Lernstrategien“ wird in Modell 4 untersucht. Die Ergebnistabelle der logistischen Multilevel-Regression weist einen signifikanten, negativen Logit-Koeffizient für die Einstellungsstärke aus. Demnach sinkt die Wahrscheinlichkeit einer DK-Antwortgabe über alle untersuchten Items mit zunehmender Einstellungsstärke (vgl. Tabelle 13, Seite 135). Der AME beschreibt die Stärke dieses negativen Zusammenhangs wie folgt: Steigt die Einstellungsstärke um eine Einheit an, verringert sich die Wahrscheinlichkeit der Wahl der DK-Kategorie um 0,3 Prozentpunkte (vgl. Tabelle 14, Seite 136).

Dieser Zusammenhang ist auch in der graphischen Darstellung der vorhergesagten Wahrscheinlichkeiten (vgl. Abbildung 31) deutlich erkennbar:

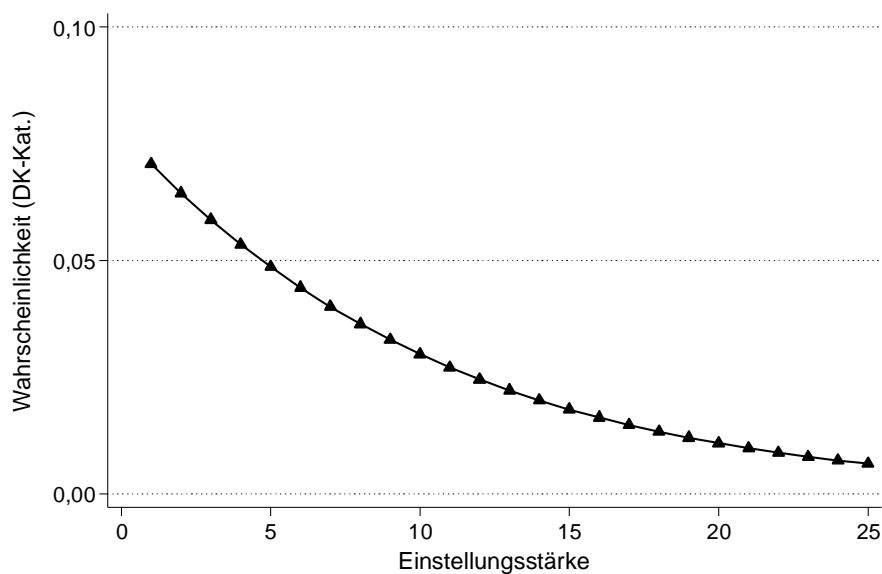


Abbildung 31 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Einstellungsstärke

Bei einer sehr niedrigen Einstellungsstärke (1 Punkt) wird eine Wahrscheinlichkeit von 7,1 % für die Wahl der DK-Kategorie vorhergesagt. Ist die Einstellungsstärke hingegen sehr hoch (25 Punkte), wird die DK-Kategorie lediglich mit einer Wahrscheinlichkeit von 1,0 % zur Antwortabgabe ausgewählt.

Für den Einfluss des Antwortkalenlayouts ist, wie im vorhergehenden Modell 3, ein signifikanter, positiver Logit-Koeffizient ausgegeben, wonach auch unter

Kontrolle der Einstellungsstärke eine niedriger Wahrscheinlichkeit einer DK-Antwortgabe in der Gruppe des ADD-DK-Layouts gegenüber dem SEP-DK-Layout bestehen bleibt (vgl. Tabelle 13, Seite 135).

Abschließend werden auch für das Modell 4 noch die weiteren Ergebnisse der logistischen Multilevel-Regression betrachtet. Die Ergebnisse blieben, im Vergleich zu dem vorangehenden Modell 3, durch die Hinzunahme der Variable „Einstellungsstärke“ unbeeinflusst. Des Weiteren ist ein Anstieg von McFaddens- R^2 von 0,085 (Modell 3) auf 0,109 (Modell 4) zu beobachten. Daraus lässt sich schließen, dass die Erklärungskraft dieses Modells durch die Hinzunahme des weiteren Prädiktors steigt. Auch die Verringerung der Werte des AIC und des BIC gegenüber den Werten des vorangehenden Modells 3 spricht dafür, dass durch die Modellerweiterung dessen Güte angehoben wird (vgl. Tabelle 13, Seite 135).

9.3.5 Modell 5: Interaktionseffekt „Einstellungsstärke × Layout“

Im letzten Modell zur Wahl der DK-Kategorie wird schließlich mittels des Interaktionsterms „Einstellungsstärke × Layout“ die Annahme der Hypothese (H 3) geprüft, dass mit steigender Einstellungsstärke die Effekte des Layouts ihre Signifikanz verlieren. Der Logit-Koeffizient für die Interaktion „SEP-DK × Einstellungsstärke“ ist nicht signifikant. Auch zeigt die Betrachtung der graphischen Darstellung der AMEs (vgl. Abbildung 32), dass sich lediglich in einem Bereich der mittleren Einstellungsstärke (12 bis 18 Punkte) geringfügig unterschiedliche Effekte auf die Wahrscheinlichkeit der Wahl der DK-Kategorie zwischen dem SEP-DK- und dem ADD-DK-Layout zeigen.

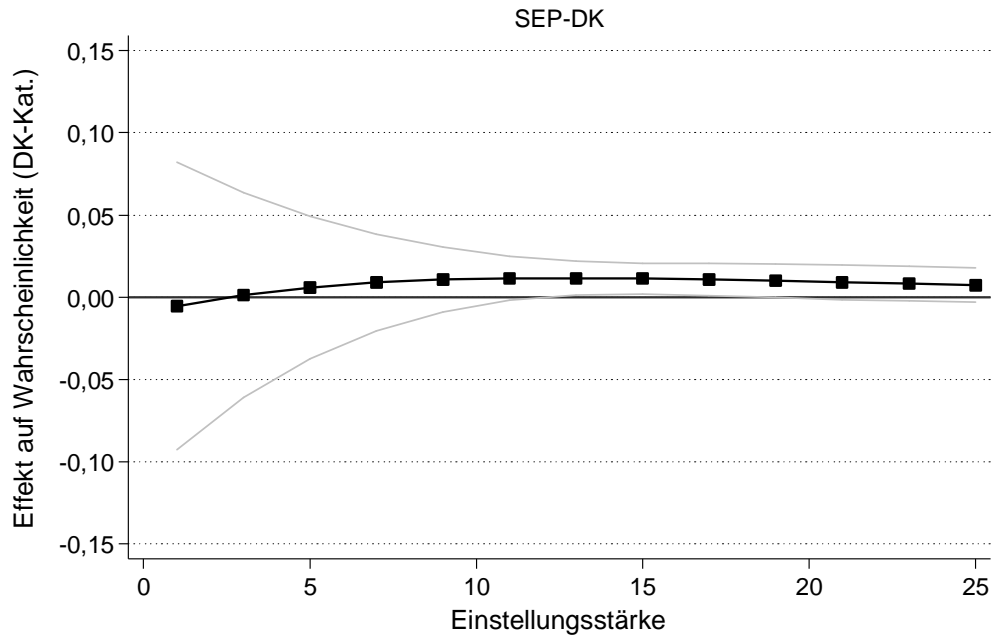


Abbildung 32 Durchschnittliche marginale Effekte (AMEs) auf die Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Einstellungsstärke

Die Betrachtung der vorhergesagten Wahrscheinlichkeiten (vgl. Abbildung 33) verdeutlicht, dass es sich um eher geringe Unterschiede handelt:

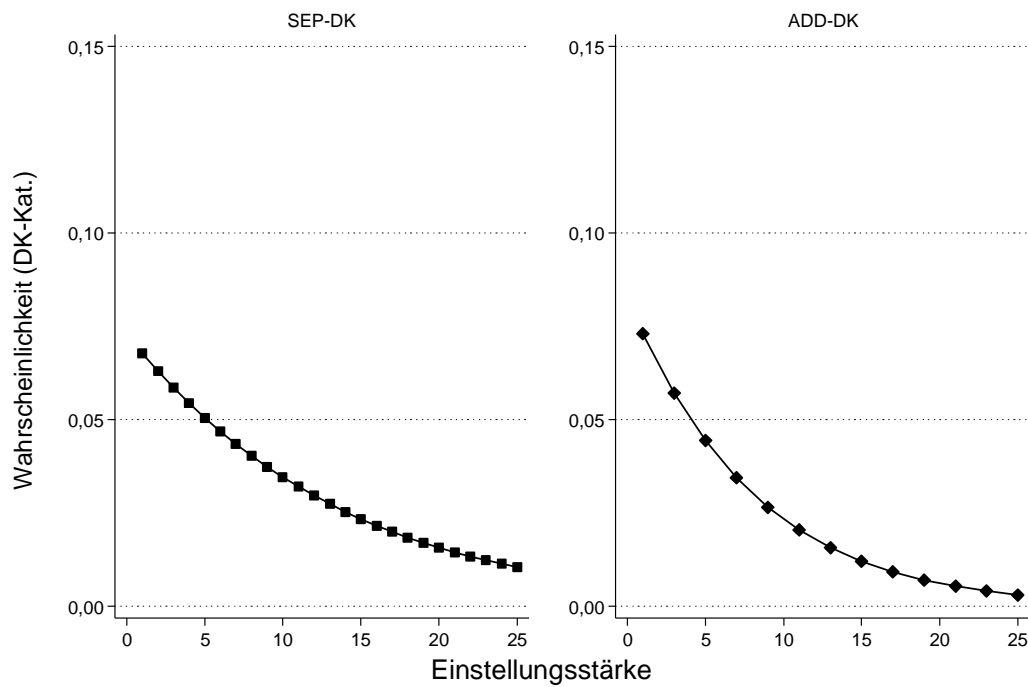


Abbildung 33 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout und Einstellungsstärke

In beiden Layoutgruppen haben Befragte mit einer niedrigen Einstellungsstärke (1 bis 5 Punkte) eine vorhergesagte Wahrscheinlichkeit der Wahl der DK-Kategorie von 5 % bis 7 %. Die Abnahme der Wahrscheinlichkeit fällt im ADD-DK-Layout insgesamt etwas stärker aus, denn hier wird für Befragte ab einer Einstellungsstärke von 14 Punkten eine Wahrscheinlichkeit von 1 % ausgewiesen. Im SEP-DK-Layout ist das erst für Befragte ab einer Einstellungsstärke von 21 Punkten der Fall. Daraus ist zu schließen, dass die hervorgehobene graphische Darstellung der DK-Kategorie im SEP-DK-Layout, wie angenommen, Befragte, auch solche die eine starke Einstellungsstärke zu dem Iteminhalt besitzen, dazu verleitet, diese Kategorie zu wählen (vgl. Abbildung 33).

Eine nach Item-Konstrukt getrennte Untersuchung des Einflusses der Einstellungsstärke auf das Auftreten von Layouteffekten (vgl. Abbildung 34) ergibt, dass nur bei Beantwortung der PES-Items im SEP-DK-Layout im Bereich einer mittleren Einstellungsstärke (13 bis 17 Punkte) ein etwas stärkerer Effekt auf die Wahrscheinlichkeit der Wahl der DK-Kategorie vorliegt als im ADD-DK-Layout. Für die LIST-Items werden über alle Ausprägungen der Einstellungsstärke hinweg keine signifikanten Unterschiede der Effekte auf die Wahrscheinlichkeit der Wahl der DK-Kategorie gefunden.

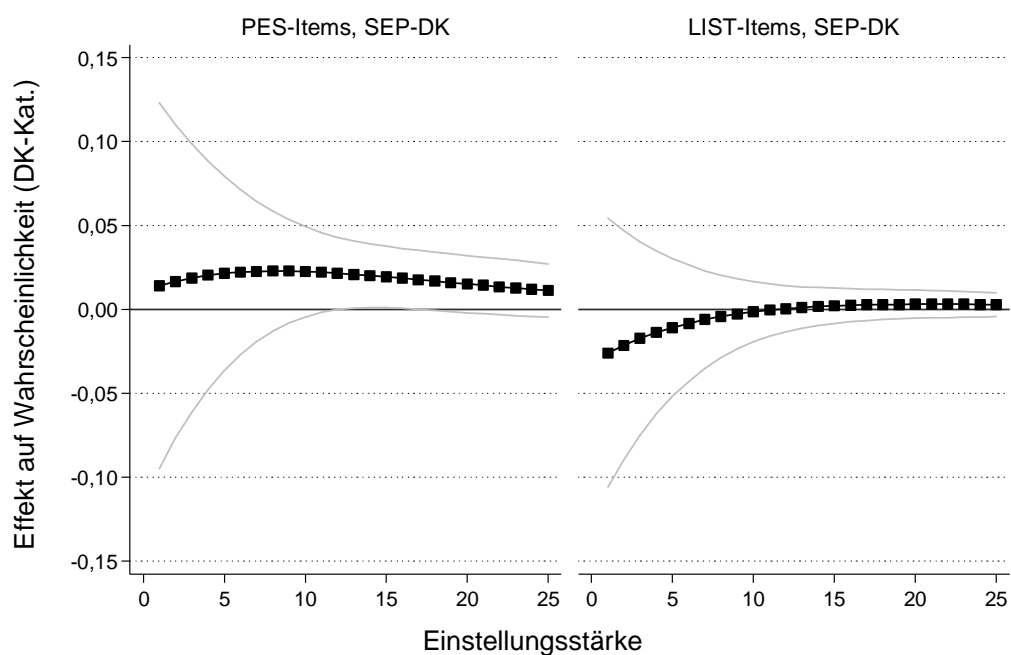


Abbildung 34 Durchschnittliche marginale Effekte (AMEs) der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK), Einstellungsstärke und Item-Konstrukt

Die Kurven der vorhergesagten Wahrscheinlichkeiten (vgl. Abbildung 35) zeigen für die PES-Items sehr ähnliche Verläufe für beide Layouts: Im Bereich der niedrigen Einstellungsstärke ist die Wahrscheinlichkeit für die Wahl der DK-Kategorie mit 9,5 % für das SEP-DK und 8,1 % für das ADD-DK-Layout recht hoch, fällt dann aber stetig ab, um im Bereich der hohen Einstellungsstärke gegen Null zu gehen. Allerdings ist der Abfall der Wahrscheinlichkeit im Bereich der unteren Einstellungsstärke im ADD-DK-Layout stärker als im SEP-DK-Layout, wie es auch bei der gemeinsamen Betrachtung aller untersuchten Items ersichtlich war. Auch die Kurven für die LIST-Items zeigen einen abfallenden Verlauf mit steigender Einstellungsstärke. Wobei im SEP-DK-Layout die Wahrscheinlichkeit von unter 5 % im niedrigsten Einstellungsbereich schon eher gering ist.

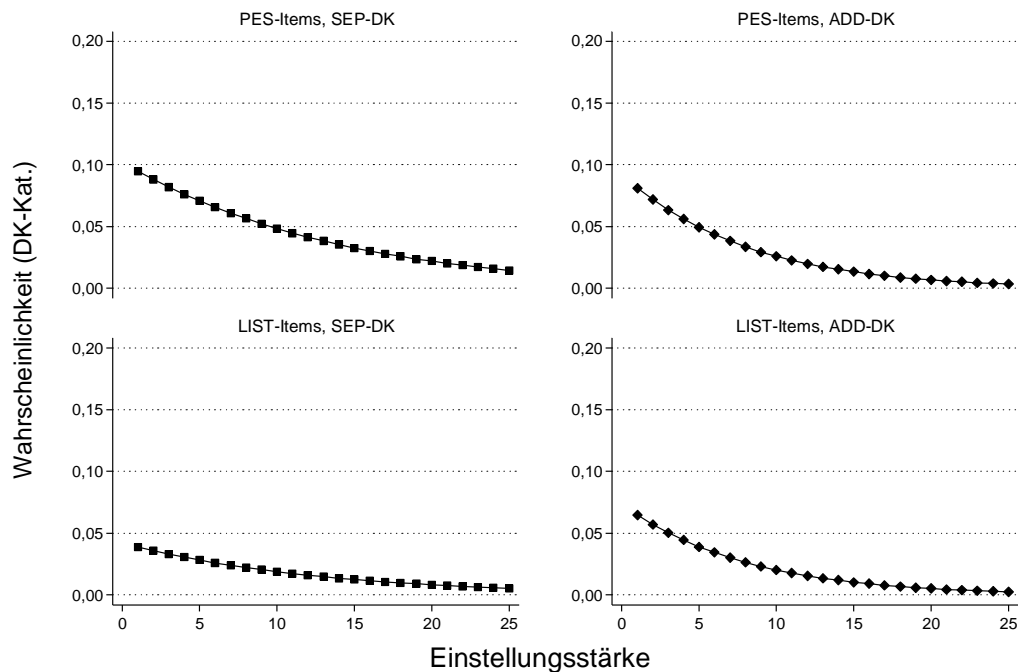


Abbildung 35 Vorhergesagte Wahrscheinlichkeiten der Wahl der DK-Kategorie nach Layout, Einstellungsstärke und Item-Konstrukt

Der Modellvergleich zeigt (vgl. Tabelle 13, Seite 135), dass in Modell 5 unter Berücksichtigung der Interaktion „Einstellungsstärke \times Layout“ der Logit-Koeffizient des SEP-DK-Layouts seine Signifikanz verliert. Die weiteren Prädiktoren des Modells 5 bleiben in ihren Effekten im Vergleich zu Modell 4 unbeeinflusst. Die Hinzunahme der Interaktion „Einstellungsstärke \times Layout“ scheint die Anpassung der Modelle an die Daten nicht zu verbessern. Weder steigt McFaddens- R^2 an, noch verringern sich die Werte des AIC und des BIC.

9.3.6 Zwischenfazit

Effekt des Layouts der Ratingskala

Die Annahme der Hypothese (H 1.3) über den Einfluss des Layouts auf die Abgabe von DK-Antworten lautet, dass durch die Abtrennung der DK-Kategorie von der Ratingskala im SEP-DK-Layout diese graphisch hervorgehoben wahrgenommen wird und dadurch in diesem Layout wahrscheinlicher gewählt wird als im ADD-DK-Layout (vgl. Unterabschnitt 5.3).

Als Ergebnis aus der Analyse des Modells 1 (vgl. Unterabschnitt 9.3.1) geht hervor, dass zwar sowohl der Logit-Koeffizient als auch der durchschnittliche marginale Effekt auf die Wahrscheinlichkeit einen Unterschied in der Wahl der DK-Kategorie in erwarteter Richtung zwischen den beiden Layouts zeigen, diese Unterschiede allerdings erst auf dem 10 % Niveau signifikant und zudem sehr gering sind. Dies unterstützt nicht die Hypothese (H 1.3).

Effekt des Item-Konstrukts

Hinsichtlich des Effekts des Item-Konstrukts besteht die Annahme der Hypothese (H 2), dass Befragte in ihrer Wahl der Antwortkategorie bei Beantwortung der Items der beiden Item-Konstrukte, aufgrund deren unterschiedlichen thematischen Inhalte und der damit verbundenen unterschiedlichen kognitiven Anforderungen, verschieden empfänglich für das Skalenlayout sind (vgl. Unterabschnitt 4.3 und Unterabschnitt 5.3).

Die Betrachtung der Wahl der DK-Kategorie zeigt, dass Befragte bei Beantwortung der PES-Items keine signifikant höhere, vorhergesagte Wahrscheinlichkeit haben, die DK-Kategorie zu wählen als bei Beantwortung der LIST-Items (vgl. Modell 2 in Unterabschnitt 9.3.2). Wird in Modell 3 (vgl. Unterabschnitt 9.3.3) die Interaktion „Item-Konstrukt \times Layout“ in der logistischen Multilevel-Regression berücksichtigt, wird erkennbar, dass der Layouteffekt Konstrukt spezifisch ist. Wie angenommen haben Befragte bei der Beantwortung der PES-Items im SEP-DK-Layout eine höhere Wahrscheinlichkeit, die DK-Kategorie zu wählen als Befragte im ADD-DK-Layout. Dieser Layouteffekt wird nicht bei Beantwortung der LIST-Items beobachtet. Demnach scheinen Befragte in der Wahl der DK-Kategorie

bestärkt zu werden, wenn diese abgetrennt von der Ratingskala dargestellt wird. Dieser Effekt tritt allerdings nur auf, wenn die Befragten zu dem abstrakteren Thema „Politische Selbstwirksamkeit“ befragt werden, nicht aber, wenn es sich um das vertrautere Thema „Lernstrategien im Studium“ handelt.

Effekt der Einstellungsstärke

Für den Einfluss der Einstellungsstärke von Befragten auf ihre Empfänglichkeit für das Antwortskalenlayout lautete die Annahme in Hypothese (H 3), dass mit steigender Einstellungsstärke Befragte weniger durch den Fragekontext beeinflusst werden und somit die Layouteffekte ihre Signifikanz verlieren. Als Resultat dessen sollten keine Unterschiede in den vorhergesagten Wahrscheinlichkeiten der Wahl der DK-Kategorie zwischen den Layoutgruppen mehr beobachtet werden können (vgl. Unterabschnitt 2.2.1 und Unterabschnitt 5.3).

Der Einbezug der Einstellungsstärke gegenüber der Politischen Selbstwirksamkeit und gegenüber den persönlichen Lernstrategien im Studium zeigt, dass die vorhergesagten Wahrscheinlichkeiten der Wahl der DK-Kategorie mit zunehmender Einstellungsstärke abnehmen (vgl. Modell 4 in Unterabschnitt 9.3.4). Dieses Ergebnis ist inhaltlich sinnvoll, denn Befragte mit einer starken Einstellung zu einem Konstrukt sollen diese mit der Wahl einer inhaltlichen Antwortkategorie zum Ausdruck bringen, anstatt die DK-Kategorie zu wählen, denn die Wahl der DK-Kategorie repräsentiert eigentlich das Fehlen einer Einstellung. Allerdings ist aus der Literatur bekannt, dass Befragte nicht nur aus Gründen von fehlenden Einstellungen die DK-Kategorie wählen, sondern auch weil sie zum Beispiel ihre Einstellung nicht von den angebotenen Antwortkategorien repräsentiert sehen, oder weil sie, aus verschiedenen Gründen, keine inhaltliche Antwort geben möchten. Daher werden auch solche Befragte mit einer starken Einstellung gegenüber den Item-Inhalten zu einem gewissen Anteil die DK-Kategorie wählen.

Die Frage, ob Befragte mit einer höheren Einstellungsstärke weniger empfänglich sind für die Effekte des Antwortskalenlayouts als Befragte mit einer niedrigen und mittleren Einstellungsstärke, wird mittels der Hinzunahme der Interaktion „Layout \times Einstellungsstärke“ beantwortet (vgl. Modell 5 in Unterabschnitt 9.3.5). Die

Ergebnisse unterstützen die Hypothese lediglich zum Teil: Für die Wahl der DK-Kategorie wird im SEP-DK-Layout für den Bereich der mittleren Einstellungsstärke ein signifikanter, positiver Effekt auf die Wahrscheinlichkeit der Wahl der DK-Kategorie ausgewiesen. Dieser Unterschied zwischen dem SEP-DK-Layout und dem ADD-DK-Layout verliert mit steigender Einstellungsstärke, wie angenommen, seine Signifikanz. Aussagen über die Effektunterschiede im Bereich der niedrigen Einstellungsstärke sind schwierig, da hier die geringen Fallzahlen hohe Standardfehler verursachen, was wiederum zu breiten Konfidenzintervallen führt.

Die Betrachtung der vorhergesagten Wahrscheinlichkeiten für die Wahl der DK-Kategorie nach Einstellungsstärke für die Gruppen SEP-DK- und ADD-DK-Layout zeigt kaum Unterschiede auf. Lediglich die stärker abfallende Steigung im ADD-DK-Layout legt den Schluss nahe, dass Befragte mit einer höheren Einstellungsstärke die DK-Kategorie aufgrund ihrer graphischen Hervorhebung im SEP-DK-Layout wählen.

Die getrennte Betrachtung der Layouteffekte nach Item-Konstrukten kommt schließlich noch zu dem Ergebnis, dass Befragte, unabhängig von ihrer Einstellungsstärke, bei Beantwortung der LIST-Items in ihrer Wahl der DK-Kategorie nicht vom Skalenlayout beeinflusst werden. Bei den PES-Items hingegen zeigen Befragte mit einer mittleren Einstellungsstärke im SEP-DK-Layout eine etwas höhere, vorhergesagte Wahrscheinlichkeit, die DK-Kategorie zu wählen als im ADD-DK-Layout.

Tabelle 13 Ergebnisse der logistischen Multilevel-Modelle zur Wahl der DK-Kategorie

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	β	β	β	β	β
	(se, robust)	(se, robust)	(se, robust)	(se, robust)	(se, robust)
Layout					
SEP-DK	0,487 (0,26)	0,488 (0,26)	0,752* (0,32)	0,791* (0,32)	0,129 (0,77)
Item-Konstrukt					
LIST		-0,720 (0,715)	-0,228 (0,76)	-0,235 (0,76)	-0,247 (0,77)
Konstrukt×Layout					
LIST×SEP-DK			-0,825 (0,55)	-0,733 (0,55)	-0,735 (0,55)
Einstellungsstärke				-0,106*** (0,03)	-0,136** (0,04)
ES⁺×Layout					
ES ⁺ ×SEP-DK					0,053 (0,06)
Konstante	-4,584*** (0,30)	-3,730*** (0,44)	-3,904*** (0,47)	-20,485*** (0,61)	-2,119** (0,69)
McFadden's-R ²	0,072	0,082	0,085	0,109	0,110
AIC	638,01	645,55	645,31	631,83	632,91
BIC	705,23	755,55	761,42	754,05	761,24
N	3331	3331	3331	3331	3331

Referenzkategorie: ADD-DK (Layout); PES (Item-Konstrukt); ⁺ES=Einstellungsstärke

* p<0,05, ** p<0,01, *** p<0,001

Tabelle 14 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der DK-Kategorie¹⁹

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)
Layout					
SEP-DK	0,009 (0,00; 0,02)	0,009 (0,00; 0,02)	0,009 (0,00; 0,02)	0,010 (0,00; 0,02)	0,011 (0,00; 0,02)
Item-Konstrukt					
LIST		-0,013 (-0,04; 0,01)	-0,013 (-0,04; 0,01)	-0,012 (-0,04; 0,01)	-0,013 (-0,04; 0,01)
Konstrukt×Layout					
PES×SEP-DK			0,018 (0,00; 0,04)	0,019 (0,00; 0,04)	0,019 (0,00; 0,04)
LIST×SEP-DK			-0,001 (-0,01; 0,01)	0,00 (-0,01; 0,01)	0,001 (-0,01; 0,01)
Einstellungsstärke				-0,003 (0,00; 0,00)	-0,002 (0,00; 0,00)
N	3331	3331	3331	3331	3331

Referenzkategorie: ADD-DK (Layout); PES-Items (Item-Konstrukt), ⁺ KI = Konfidenzintervall

¹⁹ In Modell 5 sind die AMEs zu den 25 Stufen der Einstellungsstärke nach Skalenlayout aus Gründen der Übersichtlichkeit nicht einzeln aufgeführt. Sie können jedoch den Graphen in Abbildung 32 und Abbildung 34 entnommen werden.

Tabelle 15 Vorhergesagte Wahrscheinlichkeiten (Whs.) der Wahl der DK-Kategorie²⁰

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)
Layout					
SEP-DK	0,025 (0,02; 0,03)	0,025 (0,02; 0,03)	0,025 (0,02; 0,03)	0,026 (0,02; 0,03)	0,025 (0,02; 0,03)
ADD-DK	0,016 (0,01; 0,02)	0,016 (0,01; 0,02)	0,016 (0,01; 0,02)	0,015 (0,01; 0,02)	0,015 (0,01; 0,02)
Item-Konstrukt					
PES-Items		0,027 (0,01; 0,04)	0,027 (0,01; 0,04)	0,026 (0,01; 0,04)	0,026 (0,01; 0,04)
LIST-Items		0,013 (0,00; 0,02)	0,013 (0,00; 0,02)	0,014 (0,00; 0,02)	0,013 (0,00; 0,02)

- Tabellenfortsetzung auf der nächsten Seite -

²⁰ In Modell 4 und Modell 5 werden aus Gründen der Übersichtlichkeit nicht die vorhergesagten Wahrscheinlichkeiten der 25 Ausprägungen der Einstellungsstärke sowie der Werte der Einstellungsstärke nach Antwortskalenlayout aufgeführt. Diese Werte können aus den Graphen in Abbildung 31, Abbildung 33 und Abbildung 35 entnommen werden.

- Fortsetzung der Tabelle -

	Modell 1 Whs. (95 % KI ⁺)	Modell 2 Whs. (95 % KI ⁺)	Modell 3 Whs. (95 % KI ⁺)	Modell 4 Whs. (95 % KI ⁺)	Modell 5 Whs. (95 % KI ⁺)
Konstrukt×Layout					
PES×SEP-DK			0,035 (0,01; 0,06)	0,035 (0,01; 0,06)	0,035 (0,01; 0,06)
PES×ADD-DK			0,017 (0,00; 0,03)	0,017 (0,00; 0,03)	0,016 (0,00; 0,03)
LIST×SEP-DK			0,014 (0,00; 0,03)	0,014 (0,00; 0,03)	0,014 (0,00; 0,03)
LIST×ADD-DK			0,014 (0,00; 0,03)	0,013 (0,00; 0,03)	0,013 (0,00; 0,02)
N	3331	3331	3331	3331	3331

⁺ KI = Konfidenzintervall

10. Fazit der Studie 1

Ziel der Studie 1 ist die Untersuchung, wie die graphische Darstellung der DK-Kategorie einer fünfstufigen Ratingskala die befragten Personen in ihrer Antwortabgabe beeinflusst. Darüber hinaus wird betrachtet, ob Layouteffekte unterschiedlich stark in Item-Konstrukten auftreten, die entweder ein unvertrautes / abstraktes Thema behandeln - hier die PES-Items - oder die ein vertrautes / alltägliches Thema abfragen - hier die LIST-Items. Schließlich wird in Studie 1 auch analysiert, inwieweit die persönliche Einstellungsstärke gegenüber den beiden abgefragten Themen das Auftreten von Layouteffekten modelliert. Somit wird mit Studie 1 ein Beitrag zur aktuellen Forschung geboten, die sich mit widersprüchlichen Ergebnissen hinsichtlich des Einflusses der graphischen Gestaltung von DK-Kategorien auf das Antwortverhalten von Befragten konfrontiert sieht (Christian und Dillman 2004; Tourangeau et al. 2004).

Studie 1 basiert auf Daten einer Befragung von Studenten, die selbständig Papierfragebögen ausgefüllt haben. Die Fragebögen variieren die drei interessierenden Antwortskalenlayouts „NO-DK“, „SEP-DK“ und „ADD-DK“ für die beiden Item-Konstrukte „PES“ und „LIST“. Das Antwortverhalten wird in der Analyse als drei abhängige Variablen operationalisiert: „Wahl der rechten Extremkategorie“, „Wahl der Mittelkategorie“ und „Wahl der DK-Kategorie“. Der Einfluss des Skalenlayouts auf die Wahl dieser drei Antwortkategorien wird jeweils ausführlich in einem Zwischenfazit (vgl. Unterabschnitt 9.1.6, 9.2.6 und 9.3.6) diskutiert. Daher wird im Folgenden lediglich eine Zusammenfassung der zentralen Ergebnisse der Studie 1 präsentiert.

Effekt des Layouts der Ratingskala

Hinsichtlich der Wahl der rechten Extremkategorie wird die Annahme der Hypothese (H 1.1) unterstützt, welche besagt, dass sich Befragte unter Anwendung der „middle means typical“-Heuristik bei ihrer Antwortabgabe eher am visuellen als am inhaltlichen Mittelpunkt der Ratingskala orientieren. Daher kommt es im ADD-

DK-Layout, in dem diese beiden Mittelpunkte nicht miteinander übereinstimmen, zu einer Verschiebung der Antworten hin zum rechten Skalenende. Die Analysen zeigen, dass Befragte im ADD-DK-Layout die höchste Wahrscheinlichkeit und im NO-DK-Layout die niedrigste Wahrscheinlichkeit besitzen, die rechte Extremkategorie zu wählen. Der Unterschied in der Wahrscheinlichkeit der Wahl der rechten Extremkategorie zwischen dem NO-DK- und dem SEP-DK-Layout legt des Weiteren den Schluss nahe, dass durch die angebotene, wenn auch abgetrennte DK-Kategorie die Skala im SEP-DK-Layout breiter erscheint als im NO-DK-Layout und dadurch die Extremkategorie am rechten Skalenrand in der Wahrnehmung der Befragten an Extremität verliert. Da Befragte dazu tendieren können, die Extremkategorie zu meiden (Kerlinger 1973; Tourangeau et al. 2000), kann die breitere Erscheinung der Skala im SEP-DK-Layout dazu führen, dass Befragte dort die rechte Extremkategorie wahrscheinlicher wählen als es für Befragte im NO-DK-Layout der Fall ist.

Auch hinsichtlich der Wahl der Mittelkategorie zeigen die Ergebnisse der bivariaten und multivariaten Analyse der Studie 1 ein einheitliches Bild: Die Mittelkategorie wird, wie in Hypothese (H 1.2) angenommen, von Befragten im ADD-DK-Layout deutlich weniger wahrscheinlich gewählt als von Befragten im NO-DK- oder SEP-DK-Layout. Die Annahme der Hypothese (H 1.2) ist, dass durch die klare Erkennbarkeit der Mittelkategorie im NO-DK- und im SEP-DK-Layout - hier stimmen der visuelle und der inhaltliche Mittelpunkt der Skala miteinander überein – Befragte eher zur Mittelkategorie hingezogen werden als im ADD-DK-Layout, wo diese Kategorie visuell nicht sofort erkennbar ist. Entgegen dieser Annahme zeigen sich allerdings auch signifikante Unterschiede in der Wahrscheinlichkeit der Wahl der Mittelkategorie zwischen dem NO-DK- und dem SEP-DK-Layout. Dies legt den Schluss nahe, dass neben der visuell klar erkennbaren Mittelkategorie auch die Nicht-Vorgabe einer DK-Kategorie Befragte im NO-DK-Layout am wahrscheinlichsten die Mittelkategorie wählen lässt. Aufgrund des experimentellen Designs dieser Studie ist jedoch eine weitergehende Untersuchung dieses vermuteten Effekts nicht möglich.

Hinsichtlich der Wahl der DK-Kategorie wird sowohl in der bivariaten als auch in der multivariaten Analyse deutlich, dass hier der erwartete Layouteffekt nur unter Berücksichtigung des Item-Konstrukts beobachtet wird. Für die PES-Items zeigt

sich der in Hypothese (H 1.3) formulierte Unterschied zwischen den Layoutgruppen: Befragte im SEP-DK-Layout wählen mit einer höheren Wahrscheinlichkeit die DK-Kategorie als Befragte im ADD-DK-Layout. Dieser Effekt scheint in der Betonung der DK-Kategorie durch deren graphische Abtrennung im SEP-DK-Layout begründet zu sein. Für die LIST-Items hingegen werden keine Unterschiede in der Wahl der DK-Kategorie zwischen den beiden Layoutgruppen festgestellt, was im Folgenden weitergehend diskutiert werden wird.

Effekt des Item-Konstrukts

Aus der Diskussion der konträren Ergebnisse der beiden existierenden Studien von Tourangeau et al. (2004) und Christian und Dillman (2004) über die Effekte der graphischen Darstellung der DK-Kategorie in Ratingskalen wurde die Annahme (Hypothese H 2) entwickelt, dass Layouteffekte eher dann auftreten, wenn die Items ein unvertrautes / abstraktes Thema behandeln (vgl. Unterabschnitt 4.3). Diese Annahme wird darin begründet, dass bei unvertrauten / abstrakten Fragethemen keine vorgefertigte Meinung im Gedächtnis vorliegt, weshalb sie in einem kognitiven Prozess generiert werden muss. Dabei können Befragte auf nonverbale Elemente des Fragebogens oder der Ratingskala als Unterstützung zurückgreifen. Vertraute / alltägliche Fragethemen hingegen besitzen meist eine hohe Salienz. Das heißt, sie liegen dem Befragten als schnell abrufbare Einstellung / Meinung im Gedächtnis vor und können unmittelbar daraus berichtet werden. Somit wird der kognitive Prozess stark verkürzt oder entfällt völlig, weshalb kaum ein Rückgriff auf nonverbale Elemente erwartet wird.

Diese Annahme der Hypothese (H 2) wird sowohl in der Analyse der Wahl der rechten Extremkategorie als auch der Analyse der Wahl der DK-Kategorie bestätigt. Befragte zeigen sich bei Beantwortung der PES-Items in der Wahl der jeweiligen Antwortkategorie stark durch das Skalenlayout beeinflusst, bei Beantwortung der LIST-Items hingegen wenig (Wahl der rechten Extremkategorie) bis gar nicht (Wahl der DK-Kategorie). In der Wahl der Mittelkategorie jedoch werden Befragte durch das Skalenlayout sowohl bei Beantwortung der PES- als auch der LIST-Items systematisch beeinflusst.

Aus diesen Ergebnissen lässt sich für die konträren Ergebnisse hinsichtlich des Effekts der Abtrennung der DK-Kategorie auf die Rate des Item Nonresponse der beiden publizierten Studien von Tourangeau et al. (2004) und Christian und Dillman (2004) ableiten, dass sie wahrscheinlich auf den unterschiedlichen Fragethemen (unvertraut / abstrakt vs. vertraut / alltäglich) und ihrer damit verbundenen unterschiedlichen Salienzen begründet sind. Effekte des Skalenlayouts sind vor allem dann zu erwarten, wenn Befragte einen aufwendigeren kognitiven Fragebeantwortungsprozess durchlaufen müssen, um eine Antwort zu den Items berichten zu können. In diesem Prozess scheinen die nonverbalen Elemente der Skala, die ihr Layout darstellen, von Befragten als Information gewertet und berücksichtigt zu werden.

Effekt der Einstellungsstärke

Während für die beiden Item-Konstrukte „PES“ und „LIST“ lediglich die Annahme besteht, dass es sich für Befragte dabei um ein unvertrautes / abstraktes bzw. alltägliches / vertrautes Thema handelt, wird mit der Erhebung der subjektiven Einstellungsstärke individuell für jeden Befragten erfasst, welche Wichtigkeit die beiden abgefragten Themen für ihn besitzen und wie sicher der Befragte seine Einstellung dazu berichten konnte. Die Annahme der Hypothese (H 3) ist, dass mit steigender Einstellungsstärke Befragte weniger auf nonverbale Elemente zurückgreifen und deshalb weniger durch das Skalenlayout beeinflusst werden. Diese Annahme wird von den Ergebnissen der Analyse der Studie 1 unterstützt. Bei der Wahl der rechten Extremkategorie sowie der Wahl der Mittelkategorie zeigen sich eher Befragte mit einer niedrigen bis mittleren Einstellungsstärke durch das Skalenlayout in ihrer Antwortabgabe beeinflusst. Befragte mit einer hohen Einstellungsstärke gegenüber dem abgefragten Thema wählen die rechten Extremkategorie bzw. die Mittelkategorie zur Antwortabgabe unabhängig davon, in welchem Layout die Ratingskala präsentiert ist. Auch bei der Wahl der DK-Kategorie scheint das Auftreten der Layouteffekte durch die subjektive Einstellungsstärke gegenüber dem Fragethema moderiert zu werden. Allerdings zeigen sich hier nur für den Bereich der mittleren Einstellungsstärke signifikante Layouteffekte. Eine Auswertung des Bereichs der niedrigen Einstellungsstärke ist hier aufgrund geringer Fallzahlen problematisch.

Schließlich kann aus den zentralen Ergebnissen der Studie 1 zusammenfassend geschlossen werden, dass Befragte durch das Layout der Ratingskala systematisch in ihrem Antwortverhalten beeinflusst werden. Jedoch treten Layouteffekte nicht unter allen Umständen auf, sondern hauptsächlich dann, wenn die Salienz der Einstellung niedrig ist. Bei einer niedrigen Salienz liegen die Einstellungen den Befragten nicht direkt abrufbereit im Gedächtnis vor. Vielmehr müssen sie ihre Einstellung in einem kognitiven Prozess generieren. In diesem Prozess scheinen sie die nonverbalen Elemente von Ratingskalen, also deren Layout, als Information zu bewerten und auch zu berücksichtigen. Besteht eine Diskrepanz zwischen der verbalen und der nonverbalen Information der Skala, beispielsweise hinsichtlich des Skalenmittelpunkts oder des Skalenendes, kann es zu vom Forscher ungewollten Fehlinterpretationen der Skala durch den Befragten kommen. Dies wirkt sich letztendlich auf die Qualität der Daten aus, indem beispielsweise die Rate des Item Nonresponse oder die Antworten anhand der Mittelkategorie ansteigen.

IV Studie 2

Layouteffekte in einer Online Access Panel-Befragung

11. Effekte nonverbaler Elemente in Mail und Web Survey

In dieser Dissertation werden Effekte der visuellen Darstellung der DK-Kategorie sowohl in einer selbstadministrierten Papierbefragung (Studie 1, vgl. Kapitel III) als auch in einer Online-Befragung eines Online Access Panels (Studie 2, vgl. Kapitel IV) untersucht. Bevor das Design der Studie 2 (vgl. Abschnitt 12) sowie die deskriptiven, bivariaten und multivariaten Ergebnisse (vgl. Abschnitt 13, Abschnitt 14 bzw. Abschnitt 15) vorgestellt werden, werden in diesem Abschnitt (11) zunächst kurz die Gemeinsamkeiten und Unterschiede der beiden selbstadministrierten Erhebungsmodi „Papierbefragung“ und „Online-Befragung“ umrissen. Aus dieser Gegenüberstellung der beiden Modi wird eine Annahme darüber abgeleitet, ob in Studie 1 und Studie 2 aufgrund der unterschiedlichen Erhebungsmodi unterschiedliche bzw. unterschiedlich stark ausgeprägte Layouteffekte erwartet werden. Auf Ergebnisse bestehender Untersuchungen kann nicht zurückgegriffen werden, da sich bisher keine Studie dem Vergleich von Layouteffekten in Papierfragebögen und Online-Befragungen widmet. Auch existiert bisher keine Untersuchung der Effekte des Skalenlayouts in anderen Modi als Online-Befragungen, wie beispielweise Effekte der Skalenlayouts in Papierfragebögen oder in Showcards, die in Face-to-Face-Interviews Verwendung finden.

Heutzutage erfolgt die Datenerhebung oftmals nicht mehr nur anhand eines Befragungsmodus, sondern es werden beispielweise Face-to-Face-Interviews, Telefoninterviews, postalische Befragungen und / oder Online-Befragungen in einem Umfrageprojekt miteinander kombiniert. Durch diese Kombination kann der

schwieriger werdenden Erreichbarkeit bestimmter Zielpersonen (De Leeuw et al. 2008) sowie den Präferenzen der Zielpersonen hinsichtlich der Befragungsform (Groves und Kahn 1979) Rechnung getragen werden. Im Idealfall werden damit die Ausschöpfungsquote erhöht und der Nonresponse-Bias reduziert (De Leeuw 2005; Shih und Fan 2008). Allerdings bergen diese sogenannten mixed-mode Surveys auch die Gefahr, dass Befragte je nach Befragungsmodus unterschiedlich auf dieselben Fragen antworten und somit in Abhängigkeit des Modus Messfehler entstehen (Dillman et al. 2009). Dies gefährdet die Vergleichbarkeit der mit unterschiedlichen Befragungsformen erhobenen Daten.

Ein breites Forschungsfeld setzt sich daher intensiv mit den Effekten der verschiedenen Erhebungsmodi auf die Datenqualität, das Befragtenverhalten oder die Vergleichbarkeit von mittels verschiedener Erhebungsmodi gewonnener Daten auseinander (vgl. u.a. Atkeson et al. 2014; Klausch et al. 2013; Revilla und Saris 2013). Dabei zeigt sich, dass vor allem dann Unterschiede zu beobachten sind, wenn Interviewer administrierte mit selbstadministrierten Befragungsmethoden miteinander vergleichen (Dillman et al. 2009). So tendieren zum Beispiel Befragte in Telefonumfragen stärker zur positiven Extremkategorie als Befragte in Online-Befragungen (Christian et al. 2008). Zwischen selbstadministrierten (postalischen) Papierbefragungen und Online-Befragungen zeigen sich keine, oder lediglich geringe Unterschiede in den Effekten auf die Befragten, was auf eine gute Vergleichbarkeit von Daten dieser beiden Erhebungsmethoden hindeutet (Dillman et al. 2009; Klausch et al. 2013).

Dass sich Befragte bei der Beantwortung von selbstadministrierten Papierfragebögen und von Online-Befragungen sehr ähnlich verhalten, kann durch die sich stark ähnelnden Befragungssituationen erklärt werden. In beiden Fällen handelt es sich um eine selbstadministrierte Befragung, die der Befragte ohne Unterstützung und ohne Anleitung durch einen Interviewer ausführt. Zwar wird in einem Fall das Befragungsinstrument in Papierform vorgegeben und mit dem Stift beantwortet, und im anderen Fall wird das Instrument auf dem Computer abgespielt und die Antwortabgabe wird mittels der Maus vorgenommen, die visuelle Wahrnehmung des Instruments und die damit verbundenen nonverbalen Informationen sind allerdings in beiden Modi dieselben. Auch die Anonymität der

Antwortabgabe wird sowohl in selbstadministrierten (postalischen) Papierbefragungen als auch in Online-Befragungen höher eingeschätzt als in Interviewer administrierten Modi, weshalb im selbstadministrierten Kontext Befragte weniger Effekte der sozialen Erwünschtheit zeigen und mehr wahre Antworten zu sensiblen Fragen abgeben (Tourangeau et al. 2000; Tourangeau und Yan 2007).

Ein wesentlicher Unterschied zwischen postalischen Papierbefragungen und Online-Befragungen besteht in der Feldarbeit, welche bei den postalischen Befragungen deutlich kostenintensiver und zeitaufwendiger ist als bei Online-Befragungen (Couper et al. 2001). Die Befragten selbst sind allerdings von der unterschiedlichen Administration des Surveys kaum betroffen. Lediglich bei der Rückgabe des komplettierten Fragebogens müssen Teilnehmer postalischer Befragungen die Aufgabe übernehmen, den Fragebogen in einen Briefumschlag zu stecken und bei der Post aufzugeben. Dieser Schritt entfällt bei Online-Befragungen, da die Befragungsteilnehmer am Ende der Befragung lediglich auf den „senden“-Knopf drücken müssen (Kwak und Radler 2002). Dieser etwas höhere administrative Aufwand beeinflusst allerdings die Rücklaufquote nicht negativ. Sie liegt in postalischen Befragungen im Durchschnitt sogar etwas höher als in Online-Befragungen (Shih und Fan 2008). Des Weiteren sind Online-Befragungen aufgrund ihrer automatisierten Durchführung am PC den selbstadministrierten Papierfragebögen hinsichtlich der Filterführung überlegen, denn diese läuft für die Befragten nicht sichtbar im Hintergrund ab. Dahingegen sollen kompliziertere Filterführungen in selbstadministrierten Papierfragebögen nicht eingesetzt werden, um die Belastung für den Befragten gering zu halten und um Filterfehler zu vermeiden (Porst 2009). Zudem bieten Online-Befragungen die Möglichkeit, durch interaktive Funktionen bei fehlenden oder bei DK-Antworten eine Nachfrage zu schalten. Dieses Vorgehen führt dazu, dass in Online-Befragungen im Durchschnitt die Item-Nonresponse-Rate niedriger ist als in selbstadministrierten (postalischen) Papierbefragungen (Kwak und Radler 2002).

Ein weiterer Unterschied besteht darin, dass es den Zielpersonen bei einem zugeschickten Papierfragebogen frei steht, vor dem Beantworten einzelner Fragen den ganzen Fragebogen durchzulesen, und diesen erst dann, manchmal Tage später,

auszufüllen oder unbeantwortet zu lassen. Somit ist es den Befragten möglich, lange über ihre Antworten nachzudenken, sie mit Antworten zu anderen Fragen abzustimmen oder im Nachhinein zu ändern. Außerdem nehmen Befragte von postalischen Papierbefragungen bei dieser flexiblen Handhabung des Befragungsinstruments sein gesamtes Layout wahr. Bei Online-Befragungen hingegen hängt es vom Forscher ab, ob er den Befragten solch einen Handlungsfreiraum im Umgang mit dem Befragungsinstrument gewährt. Durch ein rigideres Design des Befragungsinstruments, beispielsweise ohne die Möglichkeit zu vorherigen Seiten zurückzukehren, kann der Forscher beeinflussen, wie lange Befragte über einzelne Antworten nachdenken können und ob sie gegebene Antworten im weiteren Verlauf der Befragung abändern können. Teilnehmer von Online-Befragungen nehmen nur das jeweilige Layout einer Webseite und der darauf abgebildeten Antwortvorgaben und Antwortskalen wahr. Schließlich unterscheiden sich postalische Papierbefragungen und Online-Befragungen auch in ihrer Repräsentation der Grundgesamtheit (im Englischen Coverage), da die Selbstselektion der Zielpersonen zur Teilnahme von sozio-ökonomischen und demographischen Charakteristika bestimmt wird. Während die meisten Personen (in Nationen mit einer hohen Alphabetisierungsrate) geübt im Umgang mit Stift und Papier sind, ist ein vertrauter Umgang mit dem Internet und dem Computer nach wie vor nicht für alle Bevölkerungsgruppen selbstverständlich. So sind auch heute noch in Web Surveys meist ältere und niedrig gebildete Personen, sowie Personen, die einen niedrigen sozio-ökonomischen Standard besitzen, unterrepräsentiert (Schonlau et al. 2009). Allerdings verringerte sich dieses Problem in den vergangenen Jahren in den westlichen Nationen bereits mit der zunehmenden Computerisierung der Haushalte (Schonlau et al. 2009).

Aus dieser Gegenüberstellung zentraler Gemeinsamkeiten und Unterschiede von selbstadministrierten (postalischen) Papierbefragungen und Online-Befragungen wird die Annahme abgeleitet, dass Layouteffekte in beiden Befragungsmodi gleichermaßen beobachtbar sein sollten. In Abschnitt 1 und Abschnitt 2 dieser Dissertation wurde erläutert, dass Befragte in dem eingeschränkten konversationalen Kontext der selbstadministrierten Befragung durch das Layout der Ratingskala in ihrer Antwortabgabe beeinflusst werden, weil sie das Layout als nonverbale Information bewerten und diese Information gemäß interpretativer

Heuristiken im kognitiven Fragebeantwortungsprozess miteinbeziehen. Bei einer identischen graphischen Darstellung einer Ratingskala in einem selbstadministrierten Papierfragebogen und in einem Web Survey wird daher erwartet, dass die Befragten gleichermaßen auf die nonverbalen Elemente des Layouts als Informationsquelle zurückgreifen. Allerdings können Befragte in ihrer Empfänglichkeit für nonverbale Elemente der Skala durch verschiedene Kovariate wie „Alter“ oder „Bildung“ (vgl. Unterabschnitt 2.2.1), nach denen sich die beiden Stichproben der Studie 1 und Studie 2 voneinander unterscheiden, beeinflusst werden. In den multivariaten Analysen wird für diese Kovariate kontrolliert.

12. Studiendesign

Das wesentliche Ziel der Studie 2 ist die Untersuchung, ob die in Studie 1 beobachteten Effekte des Skalenlayouts auf das Antwortverhalten der Befragten auch im Web Modus repliziert werden können. Daher werden sowohl das Antwortverhalten wie in Studie 1 als „Wahl der rechten Extremkategorie“, als „Wahl der Mittelkategorie“ und als „Wahl der DK-Kategorie“ operationalisiert als auch die gleichen drei Skalenlayouts „NO-DK“, „SEP-DK“ und „ADD-DK“ wie in Studie 1 betrachtet. Da das Experiment der Studie 2, anders als in Studie 1, an einer heterogenen Erwachsenenstichprobe durchgeführt wurde, kann auch der Einfluss von Bildung und Alter auf das Auftreten von Layouteffekten untersucht werden.

Im folgenden Unterabschnitt (12.1) werden das Verfahren der Datengewinnung und die Stichprobenstatistik beschrieben. Unterabschnitt (12.2) stellt die Items vor, anhand welcher das Experiment durchgeführt wurde und welche die abhängigen und unabhängigen Variablen der späteren Analyse bilden. In Studie 2 wird wie in Studie 1 zur Datenanalyse das Verfahren der logistischen Multilevel-Regression mit fixen Effekten eingesetzt, was in Unterabschnitt (12.3) kurz begründet wird. In Unterabschnitt (12.4) wird schließlich der Aufbau der verschiedenen Modelle vorgestellt.

12.1 Daten

Das Experiment der Studie 2 wurde in der elften Welle des GESIS Online Access Panel Pilots erhoben. Es handelt sich dabei um eine Mehrthemenbefragung zu Geschlechterrollen, politischen Einstellungen, der Einstellung zur Zuwanderung sowie zu Persönlichkeitsfragen. In der elften Welle wurde der thematische Schwerpunkt auf den individuellen Charakter und seinen Einfluss auf die politische Meinungsbildung gelegt. Die in dieser Studie untersuchten Items zu Einstellungen gegenüber der Europäischen Union wurden zum ersten Mal in einer Befragung des GESIS Online Access Panel Pilots eingesetzt, es handelt sich demnach um keine wiederholte Messung dieser Items. Die Datenerhebung der elften Welle fand im Zeitraum vom 05.12.2012 bis zum 21.01.2013 statt und wurde von dem zuständigen Projektteam bei GESIS betreut. Allen Teilnehmern wurde am Anfang der Befragung ein Incentive in Höhe von 2 Euro für eine vollständige Befragungsteilnahme angekündigt. Von den 640 per E-Mail eingeladenen Panelisten komplettierten 493 Personen den Online-Fragebogen. Dies entspricht einer Ausschöpfungsquote von 77 %.

Von 493 Befragten sind 51,3 % männlich, 45,3 % weiblich und 3,4 % der Befragten machte keine Angabe über ihr Geschlecht. Die Befragungsteilnehmer waren zum Zeitpunkt der Befragung zwischen 18 und 78 Jahren alt; das durchschnittliche Alter der Befragten beträgt 43 Jahre ($SD = 14,91$). Die Mehrheit der Befragten, nämlich 62,7 %, besitzt einen höheren Bildungsabschluss (Fach- oder Hochschulreife), gut ein Viertel der Befragten (26,4 %) geben einen mittleren Bildungsabschluss (Mittlere Reife) an und lediglich 6,3 % der Befragten haben einen niedrigen Bildungsabschluss (Hauptschulabschluss). Schließlich haben 4,7 % der Befragten keine Angabe zu ihrem Bildungsabschluss gemacht.

Die Befragten wurden zu Beginn der Befragung zufällig einer der drei Experimentalgruppen, also den Layoutgruppen, zugeteilt. Dabei wurden 32,0 % der Befragten zur Beantwortung der Items Ratingskalen im NO-DK-Layout vorgelegt, 35,3 % der Befragten erhielten Ratingskalen im SEP-DK-Layout und 32,7 % der Befragten im ADD-DK-Layout. Die befragten Personen in den drei Layoutgruppen unterscheiden sich hinsichtlich ihres Alters ($\chi^2 = 89,16, p > 0,10$), ihrer Bildung

($\chi^2 = 1,70, p > 0,10$) oder ihres Geschlechts ($\chi^2 = 5,09, p > 0,10$) nicht signifikant.

Um Forschungsergebnisse über die Gruppe der Befragungsteilnehmer hinaus verallgemeinern zu können, muss die der Stichprobe zugrundeliegende Grundgesamtheit bekannt sein. Daher wird eine kurze Erläuterung der Stichprobenziehung des GESIS Online Access Panel Pilots vorgenommen: Zum Aufbau des Panels wurden im Frühjahr 2011 zwei zufallsbasierte Stichproben nach dem Gabler-Häder-Design (Gabler und Häder 2002; Gabler et al. 2012) gezogen, jeweils eine Stichprobe für Festnetznummern und eine für Mobilfunknummern. Die zu untersuchende Grundgesamtheit ist die Internetbevölkerung in Deutschland, Zielpersonen ohne Internet wurde dieses nicht zur Verfügung gestellt. Nach einer Screeningfrage, mit der festgestellt wurde, ob die Zielperson das Internet privat nutzt, wurde ein kurzes Interview über den Alltag und die Gesellschaft in Deutschland durchgeführt, an dessen Ende die Zielperson zu weiteren Befragungen im Internet eingeladen wurde. Dazu mussten sie ihre E-Mail-Adresse preisgeben. Da der Schwerpunkt des GESIS Online Access Panel Pilot-Projekts die Bearbeitung methodischer Fragestellungen ist, gab es in den bisherigen Wellen zahlreiche Variationen im Rekrutierungsprozess der Panelisten. So wurden beispielsweise unterschiedliche Incentivierungen der Teilnehmer für die Onlineteilnahme (von Null Euro bis zu maximal 100 Euro je Teilnehmer), aber auch Unterschiede in weichen Faktoren, z.B. Themenankündigung am Telefon und E-Mail-Einladungen, insgesamt variiert. Die Wellen werden in einem vierwöchigen Rhythmus durchgeführt und unterscheiden sich hinsichtlich ihrer Schwerpunktthemen. Die Einladung zur elften Welle ist per E-Mail erfolgt, bei Nicht-Teilnahme wurde bis zu drei Mal eine Erinnerung verschickt. Die durchschnittliche Teilnahmezeit liegt unter fünfzehn Minuten in allen bis zur elften Welle.

12.2 Items

In dem Experiment der Studie 2 werden folgende zehn Items aus der German Longitudinal Election Study (2011) eingesetzt, welche die Einstellung gegenüber der Europäischen Union erfassen (kurz EU-Items):

1. Die soziale Sicherheit in Deutschland wird durch EU-Bestimmungen geschwächt.
2. Die Regionen der EU sollen ihre Eigenständigkeit bewahren.
3. Ein Mitgliedstaat soll die EU auf eigenen Wunsch verlassen können.
4. Die Osterweiterung der EU hat zu einer wirtschaftlichen Belebung in Deutschland geführt.
5. Die Osterweiterung der EU hat die Sicherheit der Arbeitsplätze in Deutschland gefährdet.
6. Alle EU-Bürger sollen durch Volksabstimmung über EU-Verträge entscheiden können.
7. Die Osterweiterung hat zu einem Anstieg der Kriminalität in Deutschland geführt.
8. Die Einführung des Euro ist ein großer Erfolg.
9. Der Euro sollte in allen EU-Staaten eingeführt werden.
10. Die EU braucht eine gemeinsame Außen- und Sicherheitspolitik.

Die Antwortkategorien der fünfstufigen, vollverbalisierten Ratingskala sind „lehne stark ab“, „lehne eher ab“, „weder noch“, „stimme eher zu“ und „stimme stark zu“ und sind somit die gleichen wie für die PES-Items in Studie 1 (vgl. Unterabschnitt 6.2). Die DK-Kategorie ist mit „weiß nicht“ bezeichnet.

Im Gegensatz zu Studie 1, wird in Studie 2 keine Matrixdarstellung der Items gewählt, sondern eine Single-Item-Darstellung, welche in dem Screenshot des ersten EU-Items aus dem Online-Fragebogen veranschaulicht ist (vgl. Abbildung 36). In der Abbildung ist eine Ratingskala im SEP-DK-Layout zu sehen, die direkt unter dem Item-Text dargestellt ist. Jedes der zehn EU-Items wurde auf einer eigenen Seite des Online-Fragebogens abgebildet.

Die soziale Sicherheit in Deutschland wird durch EU-Bestimmungen geschwächt.					
lehne stark ab	lehne eher ab	weder noch	stimme eher zu	stimme stark zu	weiß nicht
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 36 Screenshot der Single-Item-Darstellung des ersten EU-Items in der Online-Befragung der Studie 2

12.3 Methode

Für die Analyse der Experimentaldaten der Studie 2 wird das selbe Verfahren wie in Studie 1 angewandt, da hier zum einen die gleichen dichotomen abhängigen Variablen „Wahl der rechten Extremkategorie“, „Wahl der Mittelkategorie“ und „Wahl der DK-Kategorie“ betrachtet werden und zum anderen wiederum eine hierarchische Datenstruktur (Items in Befragten) vorliegt. Daher ist auch für Studie 2 die logistische Multilevel-Regression mit fixen Effekten das geeignete Verfahren (vgl. Unterabschnitt 6.3.1 und 6.3.2). Die Interpretation der Ergebnisse der Regressionsmodelle wird wiederum anhand der vorhergesagten Wahrscheinlichkeiten und der durchschnittlichen marginalen Effekte vorgenommen, da die Bewertung der inhaltlichen Bedeutung von Effekten in logistischen Regressionen nicht direkt aus den Logit-Koeffizienten geschlossen werden kann. Dieses Vorgehen ist ausführlich in Unterabschnitt 6.3.3 erläutert.

12.4 Aufbau der logistischen Regressionsmodelle

Die multivariate Analyse der Effekte des Skalenlayouts auf die Wahl der rechten Extrem-, der Mittel- bzw. der DK-Kategorie erfolgt in zwei Schritten (vgl. Abbildung 37): In Modell 1 werden die Effekte des Layouts auf die Wahl der jeweiligen Antwortkategorie untersucht. In diesen Modellen wird für den Einfluss des Geschlechts, der Bildung und des Alters auf die Kategoriewahl kontrolliert. Im zweiten Schritt wird zum einen eine Interaktion zwischen Bildung und Layout (Modell 2a) hinzugenommen. Damit kann die Hypothese (H 4.1) überprüft werden, ob Befragte mit einem niedrigen / mittleren Bildungsabschluss aufgrund ihrer geringeren kognitiven Leistungsfähigkeit eher das Skalenlayout zur Fragebeantwortung nutzen als Befragte mit einem hohen Bildungsabschluss. Zum anderen wird in Modell 2b durch die Hinzunahme einer Interaktion zwischen Alter und Layout die Hypothese (H 4.2) überprüft, ob Befragte mit zunehmendem Alter aufgrund abnehmender kognitiver Fähigkeiten während der Fragebeantwortung empfänglicher für das Layout der Ratingskala werden (vgl. Unterabschnitt 5.3).

Abhängige Variablen	
-	Wahl der Extremkategorie
-	Wahl der Mittelkategorie
-	Wahl der DK-Kategorie
Unabhängige Variablen	
Modell 1:	Skalenlayout, Geschlecht, Bildung, Alter (vgl. Unterabschnitt 15.1)
Modell 2a:	+ Interaktion „Bildung × Layout“ (vgl. Unterabschnitt 15.2)
Modell 2b:	+ Interaktion „Alter × Layout“ (vgl. Unterabschnitt 15.3)

Abbildung 37 Aufbau der geschachtelten logistischen Multilevel-Regressionen der Wahl der rechten Extrem-, der Mittel- bzw. der DK-Kategorie

13. Deskriptive Ergebnisse

Im folgenden Abschnitt werden die Antwortverteilungen der EU-Items (Unterabschnitt 13.1) sowie die Häufigkeit der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie bei der Beantwortung der EU-Items (Unterabschnitt 13.2) vorgestellt.

13.1 Antwortverteilung der EU-Items

Mittels der Varianzanalyse werden die Mittelwerte der zehn EU-Items der drei Layoutgruppen „NO-DK“, „SEP-DK“ und „ADD-DK“ miteinander verglichen. Signifikante Unterschiede zwischen den drei Layoutgruppen werden dabei für das zweite, das dritte und das zehnte EU-Item sowie für den summierten Index aller zehn EU-Items beobachtet (vgl. Tabelle A 28)²¹. Bei allen signifikanten Mittelwertunterschieden sind jeweils der geringste Mittelwert im NO-DK-Layout und der höchste Mittelwert im ADD-DK-Layout zu beobachten.

Die Mittelwerte der EU-Items reichen im NO-DK-Layout von $M_{\text{Item 9}} = 2,91$ (SD = 1,25) bis $M_{\text{Item 10}} = 3,88$ (SD = 0,89). Im SEP-DK-Layout werden Item-Mittelwerte in Höhe von $M_{\text{Item 5}} = 3,00$ (SD = 1,09) bis $M_{\text{Item 10}} =$

²¹ Alle in der Beschriftung mit Tabelle A gekennzeichneten Tabellen befinden sich in Anhang

4,04 (SD = 0,95) beobachtet. Die Mittelwerte der Items im ADD-DK-Layout reichen schließlich von $M_{\text{Item 8}} = 3,01$ (SD = 1,34) bis $M_{\text{Item 10}} = 4,24$ (SD = 0,88). Der summierte Mittelwert über alle zehn Items beträgt im NO-DK-Layout $M = 3,32$ (SD = 0,37), im SEP-DK-Layout $M = 3,43$ (SD = 0,38) und im ADD-DK-Layout $M = 3,57$ (SD = 0,38) (vgl. Tabelle A 28).

Die Verteilungen der zehn untersuchten Items weichen in allen drei experimentellen Bedingungen signifikant von der Standardnormalverteilung ab (vgl. Tabelle A 29 bis Tabelle A 38). Der höchste Shapiro-Wilk-Wert wird für Item 1 im ADD-DK-Layout ausgewiesen (0,906, $p < 0,01$; die Werte der Schiefe der Verteilungen reichen von $S = -1,47$ (SE = 0,25) für EU-Item 10 im ADD-DK-Layout bis $S = 0,26$ (SE = 0,23) für EU-Item 5 im SEP-DK-Layout. Die Werte der Kurtosis reichen von $K = -129$ (SE = 0,45) für EU-Item 9 im SEP-DK-Layout bis $K = 2,42$ (SE = 0,49) für EU-Item 10 im ADD-DK-Layout.

Für die summierten Skalenmittelwerte über alle zehn EU-Items werden Unterschiede zwischen den Bedingungen beobachtet: Im ADD-DK-Layout folgt die Verteilung der summierten Mittelwerte einer Normalverteilung, nicht aber in den anderen beiden Layoutgruppen (vgl. Tabelle A 39).

Aus diesen deskriptiven Ergebnissen der Verteilungen der EU-Items sowie der summierten Mittelwerte sind Unterschiede zwischen den drei Layoutgruppen ersichtlich, die in der bivariaten Analyse in Abschnitt (14) und in der multivariaten Analyse in Abschnitt (15) weiter untersucht werden.

13.2 Extreme, mittlere und Don't Know-Antworten zu den EU-Items

Im Folgenden werden die Verteilungen der Antworten zu den EU-Items anhand der rechten Extrem-, der Mittel- und der DK-Kategorie betrachtet, die in der späteren multivariaten Analyse die abhängigen Variablen bilden.

Bei Beantwortung der zehn EU-Items wird die rechte Extremkategorie „stimme stark zu“ von 71,2 %, die Mittelkategorie „weder noch“ von 69,2 % und die DK-Kategorie von 31,9 % der Befragten zur Antwortabgabe mindestens eines der zehn EU-Items ausgewählt (vgl. Tabelle 16).

Die Betrachtung der Auswahlhäufigkeiten, die angeben, wie oft ein Befragter die jeweilige Kategorie zur Beantwortung der zehn EU-Items gewählt hat, zeigt, dass die meisten Befragten, die Gebrauch von der rechten Extrem- und der Mittelkategorie machten, diese für ein bis drei Items zur Antwortabgabe ausgewählt haben. Die DK-Kategorie wurde, wenn sie denn zur Antwortgabe gewählt wurde, von der Mehrheit der Befragten lediglich ein oder zweimal zur Beantwortung der EU-Items genutzt. Lediglich knapp ein Prozent der Befragten, welche die DK-Kategorie zur Antwortgabe wählten, tat dies für sieben bzw. neun der zehn EU-Items. Soweit dies überhaupt möglich ist, wird aus dieser Verteilung geschlossen, dass in den Daten kaum Antworttendenzen zu den hier betrachteten Antwortkategorien vorliegen und daher keine Fälle aus der Analyse ausgeschlossen werden müssen.

Tabelle 16 Auswahlhäufigkeiten der rechten Extrem-, Mittel- und DK-Kategorie für EU-Items

	rechte Extrem- kategorie	Mittel- kategorie	DK- Kategorie
Gewählt „nein“	28,8 %	30,8 %	68,1 %
Gewählt „ja“	71,2 %	69,2 %	31,9 %
	100 %	100 %	100 %
	(N = 493)	(N = 493)	(N = 335)
<i>wenn gewählt (=ja), dann zur Antwortabgabe von</i>			
...1 Item	31,9 %	39,0 %	59,8 %
...2 Items	23,1 %	31,4 %	21,5 %
...3 Items	22,8 %	17,3 %	9,3 %
...4 Items	11,4 %	8,8 %	3,7 %
...5 Items	7,1 %	2,3 %	3,7 %
...6 Items	3,1 %	1,2 %	-
...7 Items	0,6 %	-	0,9 %
...8 Items	-	-	-
...9 Items	-	-	0,9 %
...10 Items	-	-	-
	100 %	100 %	100 %
	(N = 351)	(N = 341)	(N = 107)

14. Bivariate Ergebnisse der Kategoriewahl in Abhängigkeit von Skalenlayout

In diesem Abschnitt wird die bivariate Analyse der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie in Abhängigkeit der drei untersuchten Skalenlayouts vorgestellt.²² Die Analyse erfolgt wie in Studie 1 auf Skalen- und nicht auf Itemebene. Daher wird von der Wahl einer Antwortkategorie gesprochen, wenn ein Befragter die untersuchte Kategorie für mindestens eines der zehn EU-Items zur Beantwortung ausgewählt hat.

Rechte Extremkategorie

In der Wahl der rechten Extremkategorie sind zwischen den drei Layoutgruppen deutliche Unterschiede zu sehen (vgl. Tabelle 17): Im ADD-DK-Layout wählen mit 82,6 % mit Abstand die meisten Befragten die rechte Extremkategorie mindestens einmal zur Beantwortung der zehn EU-Items aus. Im NO-DK-Layout sind es 60,8 % der Befragten und somit 21,8 Prozentpunkte weniger als im ADD-DK-Layout. Im SEP-DK-Layout entscheiden sich 70,1 % der Befragten für die rechte Extremkategorie und somit 11,5 Prozentpunkte weniger als im ADD-DK-Layout. Gemäß dem Chi-Quadrat-Test sind diese Unterschiede signifikant. Die unterschiedlichen Auswahlhäufigkeiten der rechten Extremkategorie in den drei Layoutgruppen unterstützen die Annahme der Hypothese (H 1.1): Weitaus mehr Befragte wählen die Extremkategorie am rechten Skalenende, wenn der graphische Skalenmittelpunkt in diese Richtung verschoben wurde und wenn die Extremkategorie visuell nicht deutlich als solche wahrnehmbar ist. Der Unterschied zwischen dem NO-DK-Layout und dem SEP-DK-Layout lässt zudem den Schluss zu, dass durch die angebotene, wenn auch abgetrennte DK-Kategorie die Skala im SEP-DK-Layout ebenfalls breiter erscheint als im NO-DK-Layout und dadurch die Extremkategorie am rechten Skalenrand in der Wahrnehmung der Befragten etwas

²² Da hier für die verschiedenen Ausprägungen der Kovariate „Bildung“ und „Alter“ nicht die prozentualen Auswahlhäufigkeiten der rechten Extrem-, der Mittel- und der DK-Kategorie von Interesse sind, wird keine bivariate Analyse dieser Zusammenhänge vorgenommen. In der Studie 2 sind lediglich die moderierenden Einflüsse dieser beiden Kovariate auf das Auftreten von Layouteffekten von Interesse, welche in den logistischen Multilevel-Regressionen in Unterabschnitt (15.2) und Unterabschnitt (15.3) tiefergehend behandelt werden.

an Extremität verliert, wodurch mehr Befragte diese Kategorie im SEP-DK- als im NO-DK-Layout wählen.

Tabelle 17 Wahl der rechten Extrem-, Mittel- und DK-Kategorie zu den EU-Items nach Skalenlayout

	Layout		
	NO-DK	SEP-DK	ADD-DK
rechte Extremkategorie	60,8 %	70,1 %	82,6 %
Mittelkategorie	77,8 %	64,4 %	65,8 %
DK-Kategorie	-	31,6 %	32,3 %
N	158	174	161

$$\chi^2_{Ex. Kat} = 18,717, df 2, p = 0,000$$

$$\chi^2_{Mit. Kat.} = 8,299, df 2, p = 0,016$$

$$\chi^2_{DK Kat.} = 0,018, df 1, p = 0,493$$

Die hier aufgeführten Prozentwerte geben an, wie viele Befragte einer Layoutgruppe die jeweilige Antwortkategorie für mindestens eines der zehn EU-Items zur Antwortabgabe gewählt haben.

Mittelkategorie

Bei dem Vergleich der Wahl der Mittelkategorie über die drei Skalenlayouts wird ersichtlich, dass diese Kategorie am häufigsten im NO-DK-Layout gewählt wird. Hier liegt die Rate bei 77,8 % der Befragten. Sowohl im ADD-DK-Layout als auch im SEP-DK-Layout nutzen rund 13 Prozentpunkte weniger Befragte diese Kategorie zur Antwortabgabe als im NO-DK-Layout (vgl. Tabelle 17). Diese Verteilung der Wahl der Mittelkategorie nach Skalenlayout unterscheidet sich gemäß dem Chi-Quadrat-Test signifikant voneinander. Diese unterschiedlichen Anteile der Wahl der Mittelkategorie in Abhängigkeit des Layouts lassen darauf schließen, dass Befragte im NO-DK-Layout die Mittelkategorie als Ersatz für die fehlende DK-Kategorie nutzen. Da keine Unterschiede zwischen den Auswahlhäufigkeiten des SEP-DK-Layouts und des ADD-DK-Layouts beobachtet werden, scheint nicht so sehr der verschobene graphische Mittelpunkt im ADD-DK-Layout Befragte in der Wahl dieser Kategorie zu beeinflussen. Demnach wird die Annahme der Hypothese (H 1.2) durch diese Daten nicht unterstützt.

DK-Kategorie

Für die letzte hier untersuchte Kategorie, die DK-Kategorie, zeigen sich keine signifikanten Unterschiede zwischen den beiden Antwortskalenlayouts. Im SEP-DK-Layout wählen 31,9 % und im ADD-DK-Layout 31,4 % der Befragten die nicht-inhaltliche Kategorie zur Beantwortung mindestens eines der zehn EU-Items (vgl. Tabelle 17). Dieses Ergebnis unterstützt nicht die in Hypothese (H 1.3) getroffene Annahme, dass durch die graphische Abtrennung der nicht-inhaltlichen Kategorie im SEP-DK-Layout diese graphisch hervorgehoben und somit häufiger von Befragten gewählt würde als im ADD-DK-Layout.

15. Ergebnisse der logistischen Multilevel-Regressionen

Dieser Abschnitt behandelt die Ergebnisse der multivariaten Analyse des Einflusses des Skalenlayouts auf das Antwortverhalten der Befragten. Dabei werden die drei betrachteten abhängigen Variablen „Wahl der rechten Extremkategorie“, „Wahl der Mittelkategorie“ bzw. „Wahl der DK-Kategorie“ wie in Studie 1 als dichotome Variablen operationalisiert, die für jedes der zehn EU-Items jeweils angeben, ob die interessierende Kategorie zur Antwortabgabe gewählt oder nicht gewählt wurde.

Der folgende erste Unterabschnitt (15.1) präsentiert die Ergebnisse des Modells 1 über den Einfluss des Layouts auf die Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie. Hierbei ist die zentrale Fragestellung, ob die in dem PAPI-Experiment mit Studenten der Studie 1 aufgetretenen Layouteffekte auch für eine heterogene Erwachsenenpopulation im Web-Modus beobachtet werden können, ob es sich also um stabile und nicht um zufällige Effekte handelt. Im Anschluss daran werden die Ergebnisse in Modell 2a über den moderierenden Einfluss der Bildung (Unterabschnitt 15.2) und des Alters in Modell 2b (Unterabschnitt 15.3) auf das Auftreten von Layouteffekten vorgestellt.

15.1 Modell 1: Effekte des Layouts

In Modell 1 wird als erklärende Variable eine nominale Variable für das Skalenlayout einbezogen, wobei das ADD-DK-Layout als Referenzkategorie gegenüber dem NO-DK-Layout und dem SEP-DK-Layout definiert wurde. Zudem werden die demographischen Variablen „Geschlechter“, „Bildung“ und „Alter“ als Kontrollvariablen in die Analyse miteinbezogen.²³ Im folgenden Unterabschnitt (15.1.1) wird der Layouteffekt auf die Wahl der rechten Extremkategorie, in Unterabschnitt (15.1.2) auf die Wahl der Mittelkategorie und in Unterabschnitt (15.1.3) auf die Wahl der DK-Kategorie untersucht.

15.1.1 Wahl der rechten Extremkategorie

Die logistische Multilevel-Regression der Wahl der rechten Extremkategorie weist sowohl für das NO-DK-Layout als auch für das SEP-DK-Layout in Referenz zum ADD-DK-Layout signifikante, negative Logit-Koeffizienten aus (vgl. Tabelle 18, Seite 166). Demnach ist die mittlere Wahrscheinlichkeit der Wahl der rechten Extremkategorie für die untersuchten Items im NO-DK-Layout und im SEP-DK-Layout signifikant niedriger als im ADD-DK-Layout. Hinsichtlich der Stärke der Layouteffekte zeigen die AMEs (vgl. Abbildung 38), dass der Effekt auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie im NO-DK-Layout signifikant um 10 Prozentpunkte und im SEP-DK-Layout signifikant um 7 Prozentpunkte geringer ist als im ADD-DK-Layout (vgl. auch Tabelle 19, Seite 167).

²³ Die unabhängige Variable „Alter“ wird als nominale Variable mit den zwei Ausprägungen „jüngere Befragte bis 55 Jahre“ und „ältere Befragte ab 55 Jahre“ und nicht als metrische Variable in die logistische Multilevel-Regression aufgenommen. Der Grund hierfür ist, dass im unteren und oberen Altersbereich geringe Fallzahlen vorliegen. Werden in Modell 2b die marginalen Effekte auf die Wahrscheinlichkeit der Wahl der untersuchten Antwortkategorien getrennt nach Alter betrachtet, führen geringe Fallzahlen für die einzelnen Altersjahre zu hohen Standardfehlern und somit zu großen Konfidenzintervallen, was die Bewertung der Signifikanz der Effektunterschiede mit einer größeren Unsicherheit verbindet. Weil es sich bei Modell 1 um ein in Modell 2b geschachteltes Modell handelt, wird auch hier bereits die nominale und nicht die metrische Altersvariable verwendet, um die Vergleichbarkeit der beiden Modelle zu gewährleisten.

Aus denselben Gründen wird bei der Variable „Bildung“ die Gruppe mit einem „niedrigen Bildungsabschluss“ aufgrund geringer Fallzahlen mit der Gruppe mit einem „mittleren Bildungsabschluss“ zusammengelegt. Somit wird in der Regression eine Bildungsvariable eingesetzt, die zwischen Befragten mit einem hohen und einem niedrigen/mittleren Bildungsabschluss unterscheidet.

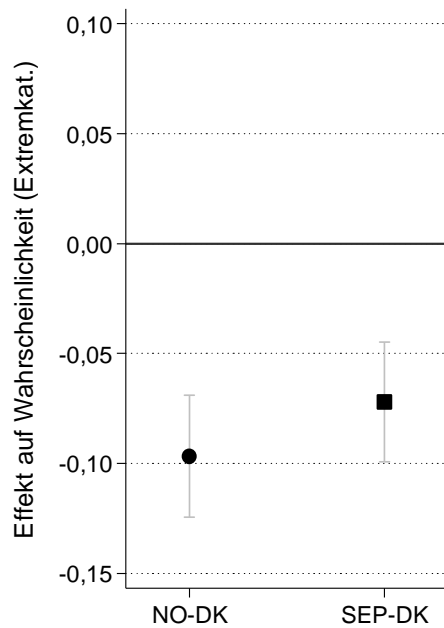


Abbildung 38 Durchschnittliche marginale Effekte (AMEs) der Wahl der rechten Extremkat. nach Layout (Referenz: ADD-DK)

Die vorhergesagten Wahrscheinlichkeiten der Wahl der rechten Extremkat. spiegeln diese Effektunterschiede zwischen den drei Layoutgruppen schließlich deutlich wider (vgl. Abbildung 39): Wie in Hypothese (H 1.1) angenommen, nutzen Befragte im ADD-DK-Layout weitaus wahrscheinlicher (23,7 %) die rechte Extremkat. zur Antwortabgabe als Befragte im NO-DK-Layout (14,1 %) und als im SEP-DK-Layout (16,5 %) (vgl. auch Tabelle 20, Seite 167).

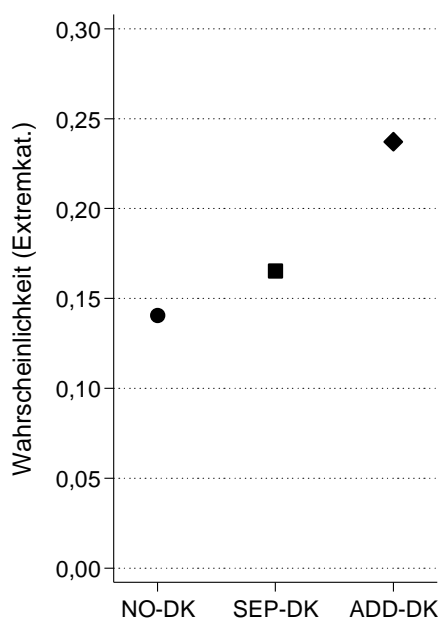


Abbildung 39 Vorhergesagte Wahrscheinlichkeiten der Wahl der rechten Extremkat.

Neben den Effekten des Skalenlayouts nehmen auch die demographischen Variablen signifikanten Einfluss auf die Wahl der rechten Extremkategorie (vgl. Tabelle 18, Seite 166): Männer wählen wahrscheinlicher als Frauen, Befragte ab 55 Jahren wahrscheinlicher als solche unter 55 Jahren und Befragte mit einem niedrigen / mittleren Bildungsabschluss wahrscheinlicher als Befragte mit einem hohen Bildungsabschluss diese Kategorie zur Antwortabgabe aus.

15.1.2 Wahl der Mittelkategorie

Für die Wahl der Mittelkategorie weist die Ergebnistabelle der logistischen Multilevel-Regression einen signifikanten, positiven Logit-Koeffizienten für das NO-DK-Layout in Referenz zum ADD-DK-Layout aus. Der Logit-Koeffizient für das SEP-DK-Layout hingegen ist nicht signifikant (vgl. Tabelle 18, Seite 166). Die AMEs geben an, dass der Effekt auf die Wahrscheinlichkeit der Wahl der Mittelkategorie im NO-DK-Layout signifikant um 4 Prozentpunkte stärker ist als im ADD-DK-Layout (vgl. Abbildung 40 und Tabelle 19, Seite 167).

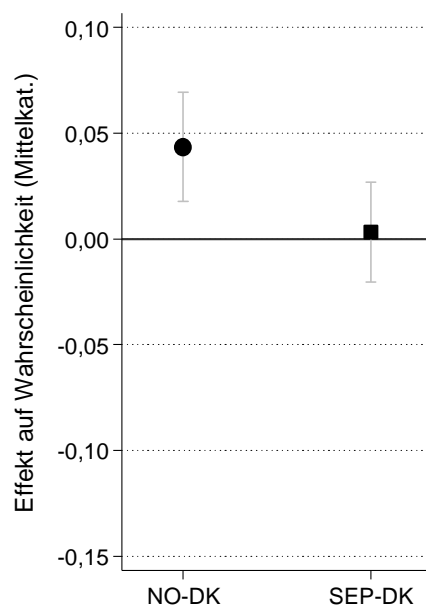


Abbildung 40 Durchschnittliche marginale Effekte (AMEs) der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK)

Der Layouteffekt spiegelt sich schließlich auch in den vorhergesagten Wahrscheinlichkeiten wider: Befragte im NO-DK-Layout haben, wie in Hypothese

(H 1.2) angenommen, mit 17,8 % eine höhere Wahrscheinlichkeit als Befragte im ADD-DK-Layout mit 13,4 % oder im SEP-DK-Layout mit 13,7 %, mindestens eines der zehn EU-Items anhand der Mittelkategorie zu beantworten (vgl. Abbildung 41 und Tabelle 20, Seite 167).

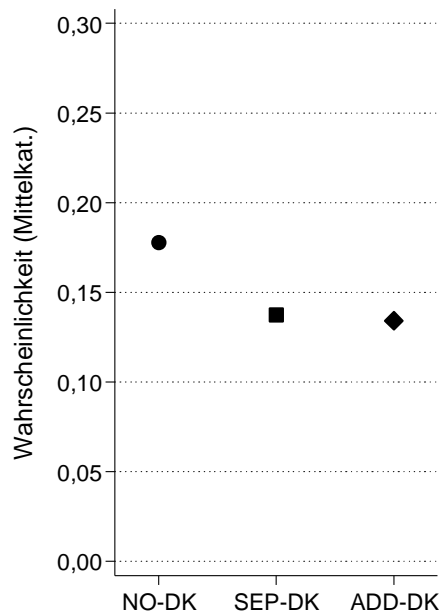


Abbildung 41 Vorhergesagte Wahrscheinlichkeit der Wahl der Mittelkategorie nach Layout

Diese Ergebnisse deuten darauf hin, dass nicht so sehr die Übereinstimmung des graphischen und des inhaltlichen Mittelpunkts, wie im NO-DK- und im SEP-DK-Layout gegeben, die Befragten die Mittelkategorie wählen lässt. Vielmehr scheint die Ursache für die hohe Wahrscheinlichkeit, im NO-DK-Layout diese Kategorie zu wählen, im Fehlen einer nicht-inhaltlichen Antwortkategorie begründet zu sein.

Des Weiteren zeigt sich, dass es hinsichtlich der Wahl der Mittelkategorie signifikante Unterschiede zwischen den demographischen Gruppen gibt (vgl. Tabelle 18, Seite 166): Befragte mit einem hohen Bildungsabschluss wählen diese Kategorie signifikant wahrscheinlicher als Befragte mit einem niedrigeren Bildungsabschluss und Befragte unter 55 Jahren wählen sie signifikant wahrscheinlicher als solche ab 55 Jahren zur Beantwortung der EU-Items aus. Signifikante Unterschiede in der Wahl der Mittelkategorie von Männern und Frauen sind hingegen nicht zu beobachten.

15.1.3 Wahl der Don't Know-Kategorie

Für die Wahl der DK-Kategorie werden schließlich in der logistischen Multilevel-Regression keine signifikanten Unterschiede zwischen den beiden Layoutgruppen SEP-DK und ADD-DK beobachtet (vgl. Tabelle 18, Seite 166).²⁴ Dies verdeutlicht auch klar die graphische Darstellung des AME, der sehr nah an der Nulllinie liegt und dessen Konfidenzintervall diese schneidet (vgl. Abbildung 42 und Tabelle 19, Seite 167). Sowohl im SEP-DK-Layout als auch im ADD-DK-Layout wählen die Befragten mit einer vorhergesagten Wahrscheinlichkeit von circa 6 % die DK-Kategorie zur Beantwortung mindestens eines der EU-Items (vgl. Tabelle 20). Daher wird mit diesen Daten des Web-Experiments nicht die Hypothese (H 1.3) unterstützt.

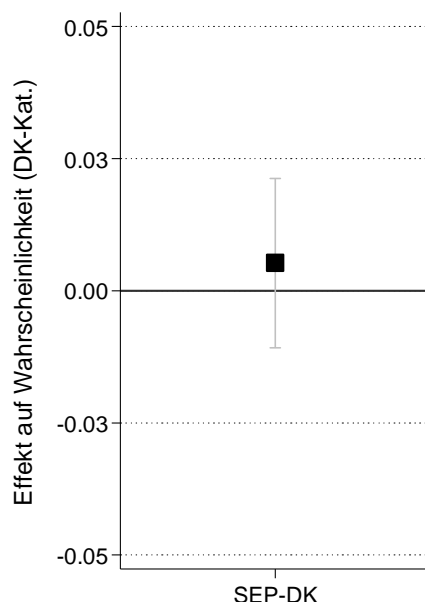


Abbildung 42 Durchschnittlicher marginaler Effekte (AME) der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK)

Des Weiteren ist der Ergebnistabelle der logistischen Multilevel-Regression zu entnehmen, dass Frauen signifikant wahrscheinlicher als Männer und Befragte unter 55 Jahren signifikant wahrscheinlicher als solche ab 55 Jahren die DK-Kategorie zur Antwortabgabe auswählen. Die Bildung nimmt keinen signifikanten Einfluss auf die Wahl der DK-Kategorie (vgl. Tabelle 18, Seite 166).

²⁴ Befragte aus der Layoutgruppe „NO-DK“ sind von der Analyse ausgeschlossen, da sie aufgrund des experimentellen Designs keine DK-Kategorie angeboten bekommen haben und somit auf der abhängigen Variable „Wahl der DK-Kategorie“ nur fehlende Werte besitzen.

15.1.4 Zwischenfazit

Die Ergebnisse der Untersuchung der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie in der Studie 2 stehen in weiten Teilen im Einklang mit den Ergebnissen der Studie 1. In Studie 1 wurde die Erkenntnis gewonnen, dass das Auftreten von Layouteffekten unter anderem abhängig von dem thematischen Item-Inhalt ist. Daher werden im folgenden Zwischenfazit die Ergebnisse zu den EU-Items der Studie 2 lediglich mit den Ergebnissen zu den PES-Items der Studie 1 verglichen, da beide Item-Konstrukte ein politisches Thema behandeln und zur Antwortabgabe für beide Item-Konstrukte die gleiche fünfstufige, vollverbalisierte Ratingskala vorgegeben wurde.

Hinsichtlich der Wahl der rechten Extremkategorie unterstützen die Ergebnisse der Studie 2 die Hypothese (H 1.1). Befragte im ADD-DK-Layout wählen um etwa 10 Prozentpunkte wahrscheinlicher die rechte Extremkategorie zur Antwortabgabe als Befragte in den beiden anderen Layoutgruppen. Damit stehen diese Ergebnisse im Einklang mit den Ergebnissen der Studie 1 (vgl. Unterabschnitt 9.1.3). Hier liegen die signifikanten Effektunterschiede bei 5 Prozentpunkten im NO-DK-Layout und bei 7 Prozentpunkten im SEP-DK-Layout, jeweils im Vergleich zu dem Effekt des ADD-DK-Layouts.

Die Ergebnisse zur Wahl der Mittelkategorie unterstützen die Hypothese (H 1.3) nur bedingt. Zwar wählen Befragte im NO-DK-Layout wie erwartet die Mittelkategorie zur Beantwortung der EU-Items wahrscheinlicher als Befragte im ADD-DK-Layout, der Effektunterschied beträgt 5 Prozentpunkte. Allerdings zeigt sich nicht der erwartete Unterschied zwischen dem SEP-DK-Layout und dem ADD-DK-Layout. Damit werden die Ergebnisse der Studie 1 auch hinsichtlich der Wahl der Mittelkategorie repliziert: In Studie 1 wird für Befragte im NO-DK-Layout bei der Beantwortung der PES-Items ein um 4 Prozentpunkte höherer Effekt auf die Wahrscheinlichkeit der Wahl der Mittelkategorie als im ADD-DK-Layout beobachtet. Die Effekte des SEP-DK-Layouts und des ADD-DK-Layouts sind auch in Studie 1 nicht signifikant unterschiedlich (vgl. 9.2.3). Damit wird die Schlussfolgerung aus den Ergebnissen der Studie 1 unterstützt, dass Befragte im NO-DK-Layout die Mittelkategorie weniger wegen ihrer visuell deutlichen Erkennbarkeit wählen, sondern vielmehr weil in diesem Layout keine DK-

Kategorie vorgegeben wird und die Mittelkategorie eine deutlich erkennbare Alternative zur DK-Kategorie bietet.

Schließlich zeigen die Ergebnisse nicht die in Hypothese (H 1.3) angenommenen Unterschiede in der Wahl der DK-Kategorie. Befragte wählen die DK-Kategorie gleichermaßen wahrscheinlich, unabhängig von ihrer graphischen Darstellung mit bzw. ohne Trennstrich. In Studie 1 werden signifikante Unterschiede zwischen dem SEP-DK-Layout und dem ADD-DK-Layout in der Wahl der DK-Kategorie beobachtet (vgl. Unterabschnitt 9.3.3).

Tabelle 18 Ergebnisse der logistischen Multilevelmodelle zur Wahl (a) der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie

	(a)	(b)	(c)
	Wahl der rechten Extremkategorie	Wahl der Mittelkategorie	DK-Kategorie ⁺
	β (se, robust)	β (se, robust)	β (se, robust)
Layout			
NO-DK	-0,690*** (0,10)	0,350*** (0,11)	-
SEP -DK	-0,487*** (0,09)	0,028 (0,11)	0,101 (0,16)
Geschlecht			
Mann	0,317*** (0,08)	-0,100 (0,09)	-0,670*** (0,16)
Bildung			
hoch	-0,244** (0,08)	0,187* (0,09)	-0,167 (0,16)
Altersgruppe			
Ältere (55 Jahre +)	0,375*** (0,09)	-0,260* (0,11)	-0,985*** (0,25)
Konstante	-2,071*** (0,19)	-1,117*** (0,15)	-1,191*** (0,18)
McFadden's-R ²	0,083	0,062	0,104
AIC	3955,01	3593,46	1269,69
BIC	4051,23	3689,68	1354,28
N	4512	4512	3109

Referenzkategorie: ADD-DK-Layout; Frau, niedrige / mittlere Bildung, Jüngere bis einschließlich 54 Jahre

* p<0,05, ** p<0,01, *** p<0,001

⁺ Von der logistischen Multilevel-Regression der Wahl der DK-Kategorie wurden alle Befragte der Layoutgruppe „NO-DK“ ausgeschlossen.

Tabelle 19 Durchschnittliche Marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl (a) der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie

	(a)	(b)	(c)
	rechte Extremkategorie	Mittelkategorie	DK-Kategorie
	AME (95 % KI ⁺)	AME (95 % KI ⁺)	AME (95 % KI ⁺)
Layout[#]			
NO-DK	-0,097 (-0,12; -,07)	0,043 (0,02; 0,07)	-
SEP-DK	-0,072 (-0,10; -0,04)	0,003 (-0,02;0,03)	0,005 (-0,01; 0,02)
N	4512	4512	3109

⁺ KI = Konfidenzintervall

[#] Referenzkategorie: ADD-DK-Layout

Tabelle 20 Vorhergesagte Wahrscheinlichkeiten (Whs.) (a) der Wahl der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie

	(a)	(b)	(c)
	rechte Extremkategorie	Mittelkategorie	DK-Kategorie
	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)	Whs. (95 % KI ⁺)
Layout			
NO-DK	0,141 (0,12, 0,16)	0,178 (0,16; 0,20)	-
SEP-DK	0,165 (0,15; 0,18)	0,137 (0,12; 0,15)	0,061 (0,05; 0,07)
ADD-DK	0,237 (0,22; 0,26)	0,134 (0,12;0,15)	0,056 (0,04; 0,07)
N	4512	4512	3109

⁺ KI = Konfidenzintervall

15.2 Modell 2a: Einfluss der Bildung

Um die Annahme der Hypothese (H 4.1) zu testen, ob die Höhe der Bildung eines Befragten moderierend auf dessen Empfänglichkeit für das Layout im Fragebeantwortungsprozess wirkt, wird das vorangehende Modell 1 um den Interaktionsterm „Bildung \times Layout“ erweitert. Die Interpretation der Effekte dieser Interaktion erfolgt wieder anhand der durchschnittlichen marginalen Effekte auf die Wahrscheinlichkeit (AMEs) der Wahl der rechten Extremkategorie (Unterabschnitt 15.2.1), der Mittelkategorie (vgl. Unterabschnitt 15.2.2) und der DK-Kategorie (Unterabschnitt 15.2.3). Schließlich werden in einem Zwischenfazit (Unterabschnitt 15.2.4) die Ergebnisse über den Einfluss der Bildung auf das Auftreten von Layouteffekten zusammengefasst und in Bezug zur Hypothese (H 4.1) (vgl. Unterabschnitt 5.3) diskutiert.²⁵

15.2.1 Wahl der rechten Extremkategorie

Die Ergebnisse der logistischen Multilevel-Regression zeigen, dass Befragte mit einem hohen Bildungsabschluss in der Wahl der rechten Extremkategorie stärker durch das Skalenlayout beeinflusst werden als Befragte mit einem niedrigen / mittleren Bildungsabschluss: Bei den höher Gebildeten ist sowohl im NO-DK-Layout als auch im SEP-DK-Layout der Effekt auf die Wahrscheinlichkeit signifikant schwächer als der Effekt des ADD-DK-Layouts. Der Effektunterschied für beide Layoutgruppen, jeweils in Referenz zum ADD-DK-Layout, beträgt rund 10 Prozentpunkte (vgl. Abbildung 43 und Tabelle 21, Seite 174). Bei Befragten mit einem niedrigen / mittleren Bildungsabschluss hingegen besteht lediglich zwischen den Layoutgruppen „NO-DK“ und „ADD-DK“ ein signifikanter Effektunterschied. Dieser ist mit 7,4 Prozentpunkten geringer als die beiden Effektunterschiede der Layoutgruppen bei Befragten mit einem hohen Bildungsabschluss (vgl. Abbildung 43 und Tabelle 21, Seite 174).

²⁵ Auf die weiteren Ergebnisse der logistischen Multilevel-Regressionsmodelle wird hier nicht mehr eingegangen, da das Hauptinteresse dieser Untersuchung auf den Interaktionseffekten der Kovariate „Bildung“ und des Skalenlayouts auf die Kategoriewahl liegt. Die Ergebnistabelle der drei logistischen Multilevel-Regressionen zur Wahl der rechten Extrem-, der Mittel- bzw. der DK-Kategorie mit der Interaktion „Bildung \times Layout“ ist im Anhang zu finden (vgl. Tabelle A 40).

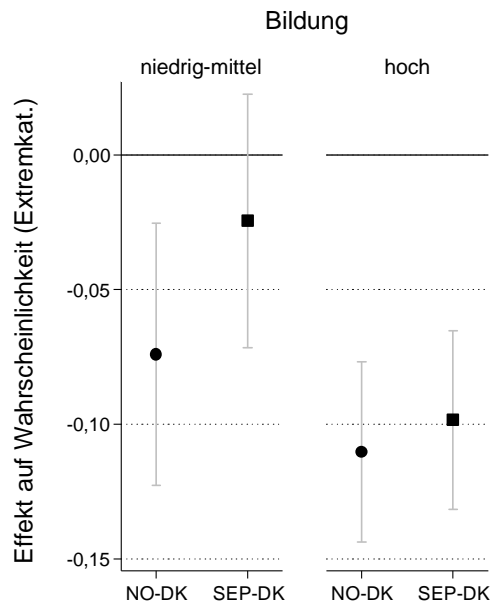


Abbildung 43 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der rechten Extremkat. nach Layout (Referenz: ADD-DK) und Bildung

15.2.2 Wahl der Mittelkategorie

Auch das Analyseergebnis der Wahl der Mittelkategorie zeigt, dass entgegen der Hypothese (H 4.1) Befragte mit einem hohen Bildungsabschluss bei der Beantwortung der zehn EU-Items empfänglicher für das Layout der Ratingskala sind. Der Effekt auf die Wahrscheinlichkeit des NO-DK-Layouts ist um 5 Prozentpunkte signifikant höher als der Effekt des ADD-DK-Layouts. Zwischen dem SEP-DK-Layout und dem ADD-DK-Layout liegt kein signifikanter Effektunterschied vor (vgl. Abbildung 44 und Tabelle 21, Seite 174). Befragte mit einem niedrigen / mittleren Bildungsabschluss hingegen zeigen sich in der Wahl der Mittelkategorie vom Skalenlayout unbeeinflusst. Hier liegen zwischen allen drei Layoutgruppen keine signifikant unterschiedlichen Effekte auf die Wahrscheinlichkeit der Wahl der Mittelkategorie vor (vgl. Abbildung 44 und Tabelle 21, Seite 174).

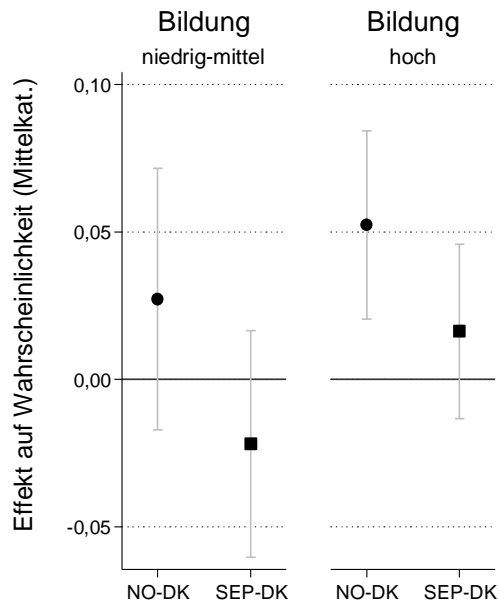


Abbildung 44 Durchschnittliche marginale Effekt (AMEs) auf die Wahrscheinlichkeit der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Bildung

15.2.3 Wahl der Don't Know-Kategorie

Die Analyse der Wahl der DK-Kategorie ohne Interaktion zwischen Bildung und Layout (Modell 1, vgl. Unterabschnitt 15.1.3) ergab, dass Befragte in beiden Layoutgruppen gleichermaßen wahrscheinlich die DK-Kategorie zur Antwortabgabe nutzen. Dieses Ergebnis bleibt auch unter Berücksichtigung eines möglichen moderierenden Einflusses der Bildung auf das Auftreten von Layouteffekten erhalten. Sowohl für Befragte mit einem niedrigen / mittleren Bildungsabschluss als auch für solche mit einem hohen Bildungsabschluss unterscheiden sich die Effekte auf die Wahrscheinlichkeit der Wahl der DK-Kategorie des SEP-DK-Layouts und des ADD-DK-Layouts nicht signifikant voneinander (vgl. Abbildung 45 und Tabelle 21, Seite 174).

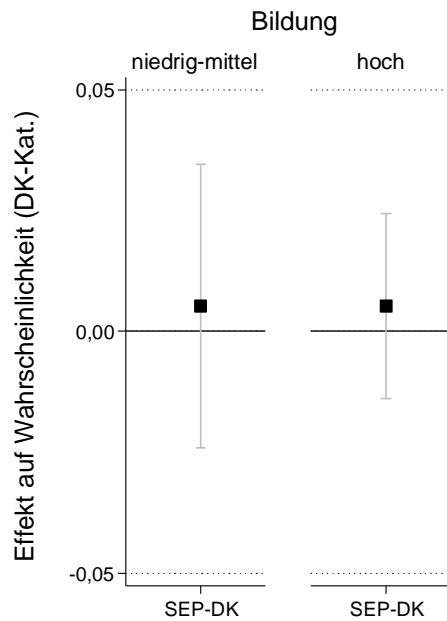


Abbildung 45 Durchschnittliche marginale Effekte (AMEs) auf Wahrscheinlichkeit der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Bildung

15.2.4 Zwischenfazit

Hinsichtlich des moderierenden Einflusses der Bildung eines Befragten wird in der Hypothese (H 4.1) die Annahme getroffen, dass Befragte mit einem hohen, im Vergleich zu Befragten mit einem niedrigen / mittleren Bildungsabschluss, weniger empfänglich für nonverbale Skalenelemente sind. Daher werden für Befragte mit hohem Bildungsabschluss geringere Layouteffekte erwartet. Begründet wird diese Annahme mit der geringeren kognitiven Leistungsfähigkeit von Befragten mit einem niedrigeren Bildungsabschluss und der damit einhergehenden stärkeren kognitiven Belastung bei der Beantwortung der EU-Items. Um die kognitive Belastung zu reduzieren, wird von niedriger gebildeten Befragten erwartet, dass sie eher ein Satisficing-Verhalten zeigen als höher gebildete Befragte, was sie wiederum empfänglicher für den Fragekontext werden lässt (Krosnick et al. 1996). Da in dieser Untersuchung Skalenlayout als Kontextinformation definiert wurde, wird angenommen, dass Befragte mit einer niedrigeren Bildung eher Layouteffekte zeigen als höher Gebildete (vgl. Unterabschnitt 2.2.1 und Unterabschnitt 5.3).

Die Ergebnisse der logistischen Regressionsmodelle der Wahl der rechten Extrem-, der Mittel- und der DK-Kategorie widerlegen diese Annahme der Hypothese (H 4.1). Befragte mit einem niedrigen / mittleren Bildungsabschluss sind in ihrer Antwortabgabe nicht stärker durch das Skalenlayout beeinflusst. Vielmehr lassen die Ergebnisse des Modells 2a darauf schließen, dass es umgekehrt, vor allem Befragte mit einem hohen Bildungsabschluss sind, die das Layout der Skala als Information bewerten und im kognitiven Fragebeantwortungsprozess nutzen. Sowohl bei der Wahl der rechten Extrem- als auch bei der Wahl der Mittelkategorie sind signifikante Effekte des Layouts zu beobachten. Bei Befragten mit einer niedrigen / mittleren Bildung zeigt sich lediglich ein Layouteffekt auf die Wahl der rechten Extremkategorie im NO-DK-Layout (in Referenz zum ADD-DK-Layout). Dieser Effekt ist allerdings schwächer als der Effekt in der Gruppe der höher Gebildeten.

Diese Ergebnisse werfen die Frage auf, warum Befragte mit einem niedrigen / mittleren Bildungsabschluss weniger bis gar nicht und Befragte mit einem hohen Bildungsabschluss hingegen stark auf das Skalenlayout reagieren? Ein Erklärungsansatz kann hier das Konzept des Kognitionsbedürfnisses (im Englischen: Need for Cognition) von Cacioppo und Petty (1982) bieten. Danach unterscheiden sich Menschen in ihrem Bedürfnis nach Kognition, also danach, in welchem Ausmaß sie anstrengende kognitive Tätigkeiten betreiben und genießen. Gemäß dieses Konzepts bilden sich Menschen mit hohem Kognitionsbedürfnis ihre eigene Meinung durch das Abwägen all der ihnen vorliegenden Argumente. Menschen mit niedrigem Kognitionsbedürfnis hingegen lassen sich oft durch begleitende Signale beeinflussen und sind in ihrem Vorgehen nicht gleichsam gründlich (Aronson et al. 2008). Es besteht die Annahme, dass sich Befragte in ihrer Strategie bei der Informationssuche und Antwortgabe im Fragebeantwortungsprozess nach ihrem Kognitionsbedürfnis unterscheiden und damit auch in ihrer Sensibilität für nonverbale Elemente. Befragte mit hohem Kognitionsbedürfnis tendieren eher dazu, alle ihnen zur Verfügung stehenden Informationen zu bewerten und für ihre Entscheidung zu Grunde zu legen als Befragte mit geringem Kognitionsbedürfnis. Deshalb sind sie auch sensibler für nonverbale Fragebogenelemente als Informationsquelle und nutzen diese eher bei der Fragebeantwortung, was auch empirisch belegt ist (Toepoel et al. 2006).

Das Kognitionsbedürfnis und die Höhe des Bildungsabschlusses haben „Intelligenz“ als gemeinsame Kovariate: Sowohl die Stärke des Kognitionsbedürfnisses (Cacioppo und Petty 1982) als auch die Wahrscheinlichkeit eines hohen Bildungsabschlusses (Deary et al. 2007) steigen mit zunehmender Intelligenz an. Daher könnte das stärkere Kognitionsbedürfnis der Befragten mit einem hohen Bildungsabschluss und die damit verbundene höhere Sensibilität für Informationen jeder Art eine Erklärung für die stärkeren Layouteffekte in dieser Befragtengruppe bieten.

Ob schließlich Befragte, die ein Satisficing-Verhalten zeigen, das Layout der Skala zur Reduzierung der kognitiven Belastung der Fragebeantwortung nutzen, lässt sich aus diesem experimentellen Design nicht schließen.

Tabelle 21 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl (a) der rechten Extrem-, (b) der Mittel- und (c) der DK-Kategorie nach Layout (Referenz: ADD-DK) und Bildung

Layout [#]	Bildung	(a)	(b)	(c)
		rechte Extremkategorie AME (95 % KI ⁺)	Mittelkategorie AME (95 % KI ⁺)	DK-Kategorie AME (95 % KI ⁺)
NO-DK	niedrig/mittel	-0,074 (-0,12; -0,03)	0,027 (-0,02; 0,07)	-
	hoch	-0,11 (-0,14; -0,08)	0,052 (0,02; 0,08)	-
SEP-DK	niedrig/mittel	-0,024 (-0,07; 0,02)	-0,022 (-0,06; 0,02)	0,005 (-0,02; 0,03)
	hoch	-0,098 (-0,13; -0,07)	0,016 (-0,01; 0,05)	0,005 (-0,01; 0,02)
<i>N</i>		4512	4512	3109

⁺KI = Konfidenzintervall

[#] Referenzkategorie: ADD-DK-Layout

15.3 Modell 2b: Einfluss des Alters

Das Modell 1 wird um den Interaktionsterm „Alter \times Layout“ erweitert, um die Annahme der Hypothese (H 4.2) zu testen, dass das Alter der Befragten als Indikator der kognitiven Leistungsfähigkeit moderierend auf das Auftreten von Layouteffekten einwirkt (vgl. Unterabschnitt 5.3).²⁶

15.3.1 Wahl der rechten Extremkategorie

Hinsichtlich der Wahl der rechten Extremkategorie zeigen sich sowohl in der Gruppe der Befragten unter 55 Jahren als auch in der Gruppe ab 55 Jahren signifikante Effekte auf die Wahrscheinlichkeit zwischen den Layoutgruppen. Insgesamt sind die Effektunterschiede zwischen dem NO-DK-Layout bzw. dem SEP-DK-Layout und dem ADD-DK-Layout bei den älteren Befragten ausgeprägter als bei den jüngeren Befragten (vgl. Abbildung 46 und Tabelle 22, Seite 179):

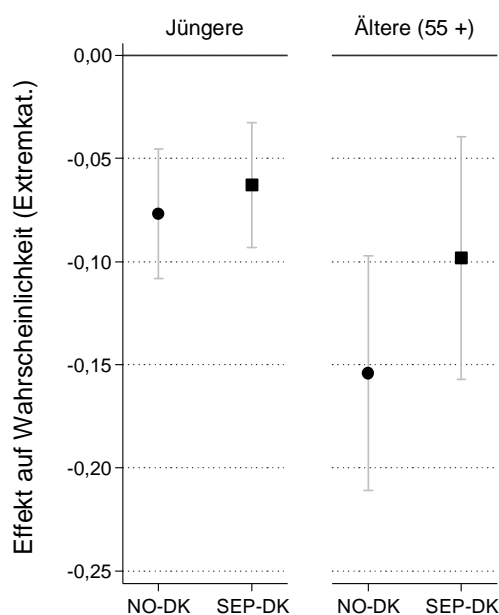


Abbildung 46 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der rechten Extremkategorie nach Layout (Referenz: ADD-DK) und Altersgruppe

²⁶ Auf die weiteren Ergebnisse der logistischen Multilevel-Regressionsmodelle wird auch hier nicht weiter eingegangen, da das Hauptinteresse dieser Untersuchung auf den Interaktionseffekten der Kovariate „Alter“ und des Skalenlayouts auf die Kategoriewahl liegt. Die Ergebnistabelle der drei logistischen Multilevel-Regressionen zur Wahl der rechten Extrem-, der Mittel- bzw. der DK-Kategorie mit Interaktion „Alter \times Layout“ ist im Anhang zu finden (vgl. Tabelle A 41).

Im NO-DK-Layout haben ältere Befragte eine um 15,4 Prozentpunkte niedrigere Wahrscheinlichkeit, die rechte Extremkategorie zu wählen als Befragte im ADD-DK-Layout. Bei den jüngeren Befragten beträgt diese Differenz zwischen den beiden Layoutgruppen mit 7,7 Prozentpunkten lediglich etwa die Hälfte. Im SEP-DK-Layout ist der Unterschied etwas schwächer. Hier haben ältere Befragte einen Effektunterschied von 9,8 Prozentpunkten und jüngere Befragte von 6,3 Prozentpunkten.

15.3.2 Wahl der Mittelkategorie

Bei der Wahl der Mittelkategorie ist für beide Altersgruppen gleichermaßen ein signifikanter Unterschied des Effekts auf die Wahrscheinlichkeit des NO-DK-Layouts und des Effekts des ADD-DK-Layouts zu erkennen. Ebenfalls ist sowohl für die jüngeren als auch für die älteren Befragten kein signifikanter Unterschied des Effekts des SEP-DK-Layouts und des ADD-DK-Layouts zu beobachten (vgl. Abbildung 47 und Tabelle 22, Seite 179). Demnach werden Befragte unabhängig ihres Alters durch das Layout der Skala beeinflusst.

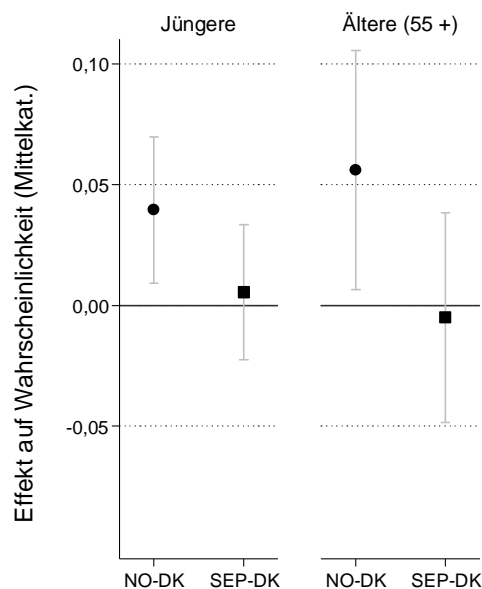


Abbildung 47 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der Mittelkategorie nach Layout (Referenz: ADD-DK) und Altersgruppe

15.3.3 Wahl der Don't Know-Kategorie

Die Effekte auf die Wahrscheinlichkeit der Wahl der DK-Kategorie sind auch unter Berücksichtigung einer möglichen Interaktion zwischen dem Alter der Befragten und dem Skalenlayout nicht signifikant unterschiedlich zwischen den Layoutgruppen (vgl. Abbildung 48 und Tabelle 22, Seite 179).

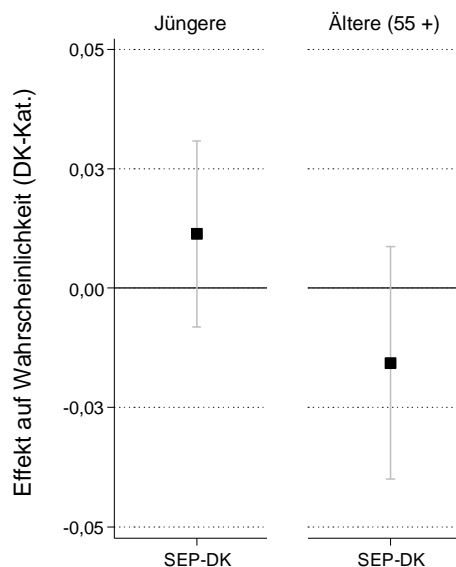


Abbildung 48 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl der DK-Kategorie nach Layout (Referenz: ADD-DK) und Altersgruppe

15.3.4 Zwischenfazit

In Hypothese (H 4.2) wird die Annahme getroffen, dass das Alter der Befragten das Auftreten von Layouteffekten moderierend beeinflusst. Begründet wird dies mit der mit steigendem Alter nachlassenden, kognitiven Leistungsfähigkeit. Darauf basierend lautet die Annahme, dass ältere Befragte einer höheren kognitiven Belastung durch den Fragebeantwortungsprozess ausgesetzt sind als jüngere Befragte, was sie wiederum zur Belastungsreduktion eher auf Kontextinformationen und somit auf nonverbale Skalenelemente zurückgreifen lässt (vgl. Unterabschnitt 2.2.1 und Unterabschnitt 5.3).

Die Ergebnisse des Modells 2b unterstützen die Hypothese (H 4.2) nicht. Sowohl bei der Wahl der rechten Extremkategorie als auch bei der Wahl der Mittelkategorie zeigen sich sowohl die jüngeren als auch die älteren Befragten gleichermaßen durch

das Skalenlayout in ihrer Antwortabgabe beeinflusst. Der Unterschied der Effekte auf die Wahrscheinlichkeit der rechten Extremkategorie fällt lediglich bei den älteren Befragten etwas stärker aus.

Tabelle 22 Durchschnittliche marginale Effekte (AMEs) auf die Wahrscheinlichkeit der Wahl (a) der rechten Extremkategorie, (b) der Mittelkategorie und (c) der DK-Kategorie nach Layout (Referenz: ADD-DK) und Altersgruppe

Layout [#]	Altersgruppe	(a)	(b)	(c)
		rechte Extremkategorie AME (95 % KI ⁺)	Mittelkategorie AME (95 % KI ⁺)	DK-Kategorie AME (95 % KI ⁺)
NO-DK	Jüngere (bis 54)	-0,077 (-0,11; -0,05)	0,040 (0,01; 0,07)	-
	Ältere (ab 55)	-0,154 (-0,21; -0,10)	0,056 (0,01; 0,11)	-
SEP-DK	Jüngere (bis 54)	-0,063 (-0,9; -0,04)	0,006 (-0,02; 0,03)	0,011 (-0,01; 0,03)
	Ältere (ab 55)	-0,098 (-0,16; -0,04)	-0,005 (-0,05; 0,04)	-0,016 (-0,04; 0,01)
<i>N</i>		4512	4512	3109

⁺KI = Konfidenzintervall

[#] Referenzkategorie: ADD-DK-Layout

16. Fazit der Studie 2

In Studie 2, wie auch in Studie 1, wird der Einfluss des Ratingskalenlayouts auf das Antwortverhalten der Befragten untersucht. Da die Teilnehmer der Online-Befragung der Studie 2 eine heterogene Stichprobe der erwachsenen Bevölkerung in Deutschland mit Internetzugang darstellen, wird die Untersuchung, ob die in der Studentenforschung der Studie 1 beobachteten Layouteffekte auch in einer heterogenen Erwachsenenstichprobe beobachtet werden können, ermöglicht. Zudem wird in Studie 2 der Einfluss von Alter und Bildung als Indikatoren für die kognitive Leistungsfähigkeit auf das Auftreten von Layouteffekten betrachtet. Unter den Teilnehmern der Online-Befragung wird die fünfstufige Ratingskala wieder randomisiert in den Layouts „NO-DK“, „SEP-DK“ und „ADD-DK“ zugeteilt. Das Antwortverhalten wird wieder als „Wahl der rechten Extremkategorie“, „Wahl der Mittelkategorie“ und „Wahl der DK-Kategorie“ operationalisiert. Die zehn untersuchten Items erfassen die „Einstellung gegenüber der Europäischen Union“, somit ein unvertrautes / abstraktes Thema. Daher werden die Ergebnisse der Studie 2 über den Einfluss des Skalenlayouts auf die Antwortabgabe zu diesen Items mit solchen Layouteffekten verglichen, die in Studie 1 für das Item-Konstrukt „Politische Selbstwirksamkeit“ (PES) – für die meisten Befragten ebenfalls ein eher abstraktes / unvertrautes Thema – beobachtet werden. Ein Unterschied im Auftreten der Layouteffekte in Studie 1 und Studie 2 aufgrund der unterschiedlichen Erhebungsmethoden wird nicht erwartet. Zwar unterscheiden sich die Papierbefragung und die Online-Befragung in einigen Punkten wesentlich voneinander, jedoch wird angenommen, dass vor allem der eingeschränkte konversationale Kontext in selbstadministrierten Befragungen, und das sind diese beiden Erhebungsmethoden gleichermaßen, die Ursache für die Beeinflussung der Befragten durch nonverbale Elemente des Fragebogens und der Ratingskala sind.

Da sowohl die Ergebnisse der Effekte des Ratingskalenlayouts auf das Antwortverhalten als auch der moderierende Einfluss des Alters und der Bildung auf das Auftreten von Layouteffekten ausführlich in Zwischenfazit direkt im Anschluss an die jeweiligen Analysen diskutiert werden, wird im Folgenden

lediglich eine Zusammenfassung der zentralen Ergebnisse der Studie 2 (im Vergleich zu denen der Studie 1) geboten.

Effekt des Layouts der Ratingskala

Den Ergebnissen der bivariaten und multivariaten Analyse der Online-Befragungsdaten der Studie 2 ist zu entnehmen, dass Befragte in ihrer Antwortabgabe durch das Skalenlayout systematisch beeinflusst werden. Sowohl für die rechte Extremkategorie als auch für die Mittelkategorie haben Befragte signifikant unterschiedliche Auswahl-Wahrscheinlichkeiten in Abhängigkeit davon, in welchem Layout ihnen die Ratingskala angeboten wird. Die beobachteten Layouteffekte stehen dabei im Großen und Ganzen im Einklang mit jenen, die in Studie 1 beobachtet wurden.

Wie in Hypothese (H 1.1) angenommen und auch in Studie 1 beobachtet, wurde die höchste Auswahl-Wahrscheinlichkeit für das ADD-DK-Layout hinsichtlich der Wahl der rechten Extremkategorie berechnet. Dies lässt darauf schließen, dass sich Befragte bei der Interpretation der Skala und bei ihrer Antwortabgabe am visuellen Skalenmittelpunkt orientieren, was im ADD-DK-Layout zu einer Verschiebung der Antworten hin zum rechten Skalenende führt.

Hinsichtlich der Wahl der Mittelkategorie wird die Hypothese (H 1.2.) auch in Studie 2 nur bedingt unterstützt: Zwar wählen Befragte im ADD-DK-Layout weniger wahrscheinlich die Mittelkategorie zur Antwortabgabe aus als Befragte im NO-DK-Layout, sie unterscheiden sich jedoch nicht in ihrer Auswahl-Wahrscheinlichkeit von den Befragten im SEP-DK-Layout. Dies spricht gegen die Annahme der Hypothese (H 1.2), nach der die Übereinstimmung des visuellen und inhaltlichen Skalenmittelpunktes im NO-DK- und im SEP-DK-Layout zu einer wahrscheinlicheren Wahl der Mittelkategorie führt, im Vergleich zum ADD-DK-Layout, in dem diese beiden Mittelpunkte voneinander abweichen. Der Unterschied in der Wahrscheinlichkeit der Wahl der Mittelkategorie im NO-DK-Layout auf der einen und im SEP-DK- und ADD-DK-Layout auf der anderen Seite legt daher den Schluss nahe, dass vor allem die Nicht-Vorgabe der DK-Kategorie im NO-DK-Layout zu einer wahrscheinlicheren Wahl der Mittelkategorie durch die Befragten führt. Sie scheinen die Mittelkategorie teilweise als Ersatz für die nicht angebotene

DK-Kategorie zu nutzen. Auch in der Studie 1 waren ähnliche Ergebnisse zu beobachten, die diesen Schluss nahelegen. Allerdings war in Studie 1 zusätzlich zu dem signifikanten Unterschied zwischen NO-DK- und SEP-DK-Layout noch ein signifikanter Unterschied zwischen dem SEP-DK- und dem ADD-DK-Layout zu sehen. In Kombination mit den Ergebnissen der Studie 2 erhärtet sich dieser Schluss, auch wenn das experimentelle Design beider Studien eine sichere Aussage darüber, ob Befragte bei der Nicht-Vorgabe von DK-Kategorien auf die Mittelkategorie ausweichen, nicht zulässt.

Schließlich werden in Studie 2 hinsichtlich der Wahl der DK-Kategorie keine Effekte des Skalenlayouts gefunden. Hier wählen Befragte diese Kategorie zur Antwortabgabe unabhängig davon aus, in welchem Layout sie die Ratingskala angeboten bekommen. Erwartet wurde in Hypothese (H 1.3), dass die graphische Abtrennung der DK-Kategorie im SEP-DK-Layout zu einer höheren Rate an Item Nonresponse führt, was in Studie 2, anders als in Studie 1, mit den Daten nicht unterstützt werden kann.

Effekte des Alters und der Bildung

Die Annahme der Hypothese (H 4.1) besagt, dass Befragte mit einer niedrigen / mittleren Bildung aufgrund ihrer geringeren kognitiven Leistungsfähigkeit vermehrt auf nonverbale Elemente des Fragekontextes zurückgreifen, um die kognitive Belastung zu vermindern. Deshalb wird erwartet, dass in dieser Befragtengruppe vermehrt Layouteffekte beobachten werden. Die Ergebnisse der Studie 2 hingegen unterstützen diese Hypothese (H 4.1) nicht. Vielmehr scheint ein umgekehrter Zusammenhang vorzuliegen: Vor allem jene Befragte mit einer hohen Bildung scheinen das Layout der Ratingskala als Information zu bewerten und beziehen es in den Fragebeantwortungsprozess mit ein. Eine Erklärung hierfür kann das Konzept des Kognitionsbedürfnisses bieten, was im Zwischenfazit (vgl. Unterabschnitt 15.2.4) ausführlicher diskutiert ist.

In Studie 2 treten Layouteffekte gleichermaßen für Befragte bis 55 Jahren und für Befragte über 55 Jahren auf. Damit wird in Studie 2 nicht die Hypothese (H 4.2) unterstützt, nach der ältere Befragte aufgrund ihrer abnehmenden kognitiven

Leistungsfähigkeit eher auf nonverbale Elemente des Fragebogens zurückgreifen, um die Belastung zu reduzieren.

Zusammenfassend kann für die Hypothesen hinsichtlich des Einflusses des Skalenlayouts auf das Antwortverhalten aus den Ergebnissen der Studie 1 und Studie 2 folgendes Fazit gezogen werden (vgl. Abbildung 49): Anhand der heterogenen Erwachsenenstichprobe der Online-Befragung werden für zwei der drei untersuchten Antwortverhalten, nämlich für die „Wahl der rechten Extremkategorie“ sowie für die „Wahl der Mittelkategorie“, sehr ähnliche Ergebnisse beobachtet wie in der Studentenbefragung der Studie 1. Dabei wird die Hypothese (H 1.1) hinsichtlich der vermehrten Wahl der rechten Extremkategorie im ADD-DK-Layout in beiden Studien unterstützt.

<i>Hypothese</i>	Studie 1	Studie 2	Übereinstimmung der Ergebnisse?
Wahl der rechten Extremkategorie (H 1.1)	✓	✓	✓
Wahl der Mittelkategorie (H 1.2)	-	-	✓
Wahl der DK-Kategorie (H 1.3)	✓	-	-

Abbildung 49 Übersicht über die Unterstützung der Hypothesen hinsichtlich des Einflusses des Skalenlayouts auf das Antwortverhalten der Ergebnisse der Studien 1 und 2 und deren Übereinstimmung

Die Hypothese (H 1.2) hinsichtlich der Wahl der Mittelkategorie erfährt hingegen in beiden Studien nur eine bedingte Unterstützung: Zwar wählen Befragte im NO-DK-Layout am wahrscheinlichsten die Mittelkategorie, der signifikante und nicht erwartete Unterschied zwischen NO-DK- und SEP-DK-Layout hingegen lässt die Annahme aufkommen, dass die Wahl der Mittelkategorie im NO-DK-Layout stärker durch die Nicht-Vorgabe einer DK-Kategorie verursacht wird als von der Tatsache, dass der visuelle und inhaltliche Mittelpunkt miteinander übereinstimmen. Besonders das Ergebnis der Studie 2, dass entgegen der Annahme kein signifikanter Unterschied in der Wahl der Mittelkategorie zwischen dem SEP-DK- und dem ADD-DK-Layout vorliegt, spricht für diese Annahme. Schließlich liefern beide Studien uneinheitliche Befunde hinsichtlich der Wahl der DK-Kategorie in Abhängigkeit vom Skalenlayout, die so nicht erwartet wurden. Wie zuvor erwähnt

besteht bei der Analyse dieser abhängigen Variablen das Problem, dass eventuell die Power des Tests nicht ausreichend ist, um diese Effekte zu beobachten. Die Erwartung, dass die Layouteffekte beider Studien trotz unterschiedlicher Erhebungsmodi ähnlich ausfallen, wird bestätigt.

V Zusammenfassende Schlussbetrachtung

Das Layout von Ratingskalen beeinflusst systematisch das Antwortverhalten in selbstadministrierten Umfragen. Für die DK-Kategorie („weiß nicht“, „kann ich nicht sagen“, „don’t know“) konnte das im Rahmen dieser Dissertation nachgewiesen werden. Allerdings moderieren Befragungskontext und Merkmale der Befragten beides - die Stärke der Layouteffekte sowie die Antwortverteilungen. Zusammengefasst ist dies das zentrale Ergebnis der beiden empirischen Studien dieser Arbeit.

Damit wird ein Beitrag zu bislang strittigen Fragen in der aktuellen Forschung geleistet, in der widersprüchliche Ergebnisse zu Layouteffekten der DK-Kategorie berichtet, aber nicht hinreichend erklärt werden konnten (Christian et al. 2009; Tourangeau et al. 2004). Widersprüchliche Ergebnisse sind für die Praxis der empirischen Forschung problematisch, weil keine konkreten Gestaltungsempfehlungen (sogenannte Best Practice) für das Layout von Ratingskalen und deren DK-Kategorie ausgesprochen werden können.

Ziel der Dissertation ist, die widersprüchlichen Ergebnisse zu erklären und darauf aufbauend praktische Empfehlungen für das Layout von Ratingskalen zu formulieren.

Dazu wurden zwei Leitfragen formuliert:

1. Beeinflusst das Layout der DK-Kategorie einer Ratingskala in selbstadministrierten Umfragen das Antwortverhalten der Befragten?
2. Wenn ja, werden entsprechende Layouteffekte durch persönliche und kontextuelle Merkmale moderiert?

In zwei empirischen Studien wurde diesen Fragen nachgegangen. Deren zentrale Ergebnisse werden hier entsprechend den in Unterabschnitt (5.3) formulierten Hypothesen zusammengefasst.

Die ersten drei Hypothesen (H 1.1, H 1.2 und H 1.3) treffen Annahmen über den Einfluss der untersuchten Ratingskalenlayouts auf das Antwortverhalten der Befragten. Diese drei betrachteten Layouts sind

1. NO-DK-Layout: eine Ratingskala ohne DK-Kategorie,
2. SEP-DK-Layout: eine Ratingskala mit einer, mittels Trennstrich abgetrennten DK-Kategorie und
3. ADD-DK-Layout: eine Ratingskala mit einer DK-Kategorie, die nicht graphisch von den inhaltlichen Antwortkategorien abgetrennt ist (vgl. auch Abbildung 51)

Das Antwortverhalten wird als „Wahl der rechten Extremkategorie“, „Wahl der Mittelkategorie“ bzw. „Wahl der DK-Kategorie“ (über alle untersuchten Items) beobachtet.

Die rechte Extremkategorie wählen Befragte, denen die Ratingskala im ADD-DK-Layout zur Antwortabgabe präsentiert wird signifikant häufiger als Befragte, die anhand der beiden anderen Layouts ihre Antwort abgeben (vgl. für Studie 1 Unterabschnitt 9.1.1, für Studie 2 Unterabschnitt 15.1.1). Dies deutet auf die Verwendung der „middle means typical“-Heuristik hin. Nach dieser Heuristik orientieren sich Befragte als Mechanismus bei der Antwortformulierung am visuellen Mittelpunkt der Skala. Im ADD-DK-Layout ist der Mittelpunkt durch die angehängte DK-Kategorie hin zum rechten Skalenende verschoben, was als ursächlich für die Verschiebung der Antworten angesehen werden kann. Beide empirische Studien, die Studentenforschung mittels selbstausgefüllten Papierfragebögen (Studie 1) und die Online-Befragung des GESIS Online Access Panel Pilots (Studie 2), stützen dieses Ergebnis (vgl. Abbildung 50).

In Studie 1 wird der Effekt des verschobenen visuellen Mittelpunkts zudem durch die Vorgabe einer DK-Kategorie im SEP-DK-Layout bzw. die Nicht-Vorgabe einer DK-Kategorie im NO-DK-Layout verstärkt: Obwohl im NO-DK-Layout und im SEP-DK-Layout eine Übereinstimmung zwischen visuellem und inhaltlichem Mittelpunkt besteht, wählen Befragte im NO-DK-Layout weniger häufig die rechte Extremkategorie aus.

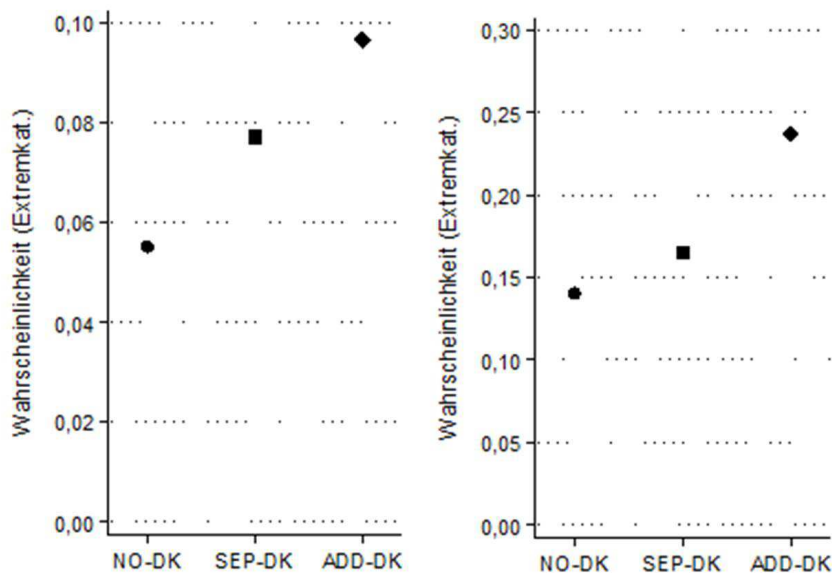


Abbildung 50 Vorhergesagte Wahrscheinlichkeit der Wahl der rechten Extremkategorie nach Layout in Studie 1 (links) und Studie 2 (rechts)

Dies kann darin begründet sein, dass im NO-DK-Layout die rechte Extremkategorie durch ihre Lage am äußeren Rand extremer erscheint als im SEP-DK-Layout, in dem sie noch von einer DK-Kategorie „eingepackt“ ist (vgl. Abbildung 51). Da Befragte die Tendenz besitzen können, die äußeren Extremkategorien zu vermeiden (Kerlinger 1973; Tourangeau et al. 2000), könnte dies eine Erklärung für die geringe Wahlwahrscheinlichkeit dieser Antwortkategorie im NO-DK-Layout darstellen.

rechte Extremkategorie					
NO-DK-Layout					
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<div>← liegt am Außenrand</div> <div>← klar erkennbar wegen fehlender DK</div>					
SEP-DK-Layout					
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	Kann ich nicht sagen
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<div>← liegt nicht am Außenrand</div> <div>← klar erkennbar wegen Abtrennung der DK</div>					
ADD-DK-Layout					
lehne stark ab	lehne ab	weder noch	stimme zu	stimme stark zu	Kann ich nicht sagen
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<div>← liegt nicht am Außenrand</div> <div>← nicht klar erkennbar wegen fehlender Abtrennung der DK</div>					

Abbildung 51 Unterschiede der Skalenlayouts in der Darstellung der rechten Extremkategorie

Die Mittelkategorie wird im NO-DK-Layout am häufigsten zur Beantwortung der Items ausgewählt. Geringe (Studie 1, vgl. Unterabschnitt 9.2.1) oder keine Unterschiede (Studie 2, vgl. Unterabschnitt 15.1.2) in der Wahl der Mittelkategorie sind hingegen zwischen dem SEP-DK-Layout und dem ADD-DK-Layout zu finden (vgl. Abbildung 52).

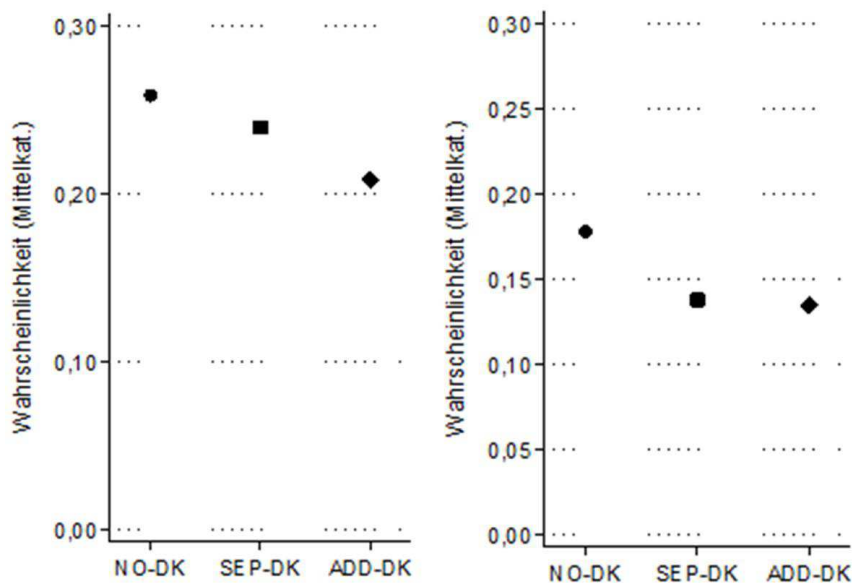


Abbildung 52 Wahl der Mittelkategorie nach Layout in Studie 1 (links) und Studie 2 (rechts)

Dieser Befund entspricht daher nur teilweise der Annahme der Hypothese (H 1.2), die besagt, dass im ADD-DK-Layout die Mittelkategorie weniger wahrscheinlich gewählt wird als in den anderen beiden Layouts, da sie aufgrund des verschobenen visuellen Mittelpunktes als Mittelkategorie für die befragten Personen nicht schnell erkenntlich ist. Eher scheint die hohe Wahrscheinlichkeit für eine Antwortabgabe anhand der Mittelkategorie im NO-DK-Layout durch die Kombination einer nicht vorgegebenen DK-Kategorie mit einer visuell klar erkennbaren Mitte zustande zu kommen. Befragte, die keine Einstellung zu dem Thema besitzen, die keine inhaltliche Antwort (aus welchen Gründen auch immer) geben möchten oder die ein Satisficing-Verhalten zeigen, können in diesem Ratingskalenlayout nicht auf eine DK-Kategorie ausweichen. In diesem Kontext gewinnt eine klar ersichtliche Mittelkategorie im NO-DK-Layout an Attraktivität gegenüber einem Layout mit DK-Kategorie. Eine genauere Aussage über die Wechselwirkung der Vorgabe von DK-Kategorien und von Mittelkategorien kann hier nicht getroffen werden, da dies das eingesetzte Studiendesign nicht zulässt. Auch der Forschungsstand lässt hierzu keine klare Aussage zu.

Entgegen der Erwartung der Hypothese (H 1.3) unterliegt die Wahl der DK-Kategorie in beiden empirischen Studien dieser Dissertation keinem Layouteffekt (vgl. für Studie 1 Unterabschnitt 9.3.1 und für Studie 2 Unterabschnitt 15.1.3), was für Studie 1 im linken Graphen der Abbildung 53 zu sehen ist. Wird allerdings in Studie 1 die Analyse getrennt nach den beiden Item-Konstrukten „Politische Selbstwirksamkeit“ (PES) und „Lernstrategien im Studium“ (LIST) vorgenommen, beeinflusst das Layout die Befragten bei der Beantwortung der PES-Items in der Wahl der DK-Kategorie in erwarteter Weise: Befragte im SEP-DK-Layout wählen häufiger die DK-Kategorie zur Antwortabgabe aus als solche im ADD-DK-Layout, was durch die Betonung dieser Kategorie durch ihre graphische Abtrennung von der Ratingskala verursacht wird. Für die Items des LIST-Inventars hingegen besteht kein Unterschied zwischen den beiden Layouts (vgl. rechter Graph der Abbildung 53 und Unterabschnitt 9.3.3). Dieses Ergebnis entspricht den Erwartungen der Hypothese (H 2), dass Layouteffekte eher bei unvertrauten, abstrakten Themen auftreten, zu denen Befragte keine direkt abrufbare Einstellung / Information im Gedächtnis besitzen, sondern die sie in einem aufwendigeren kognitiven Prozess generieren müssen.

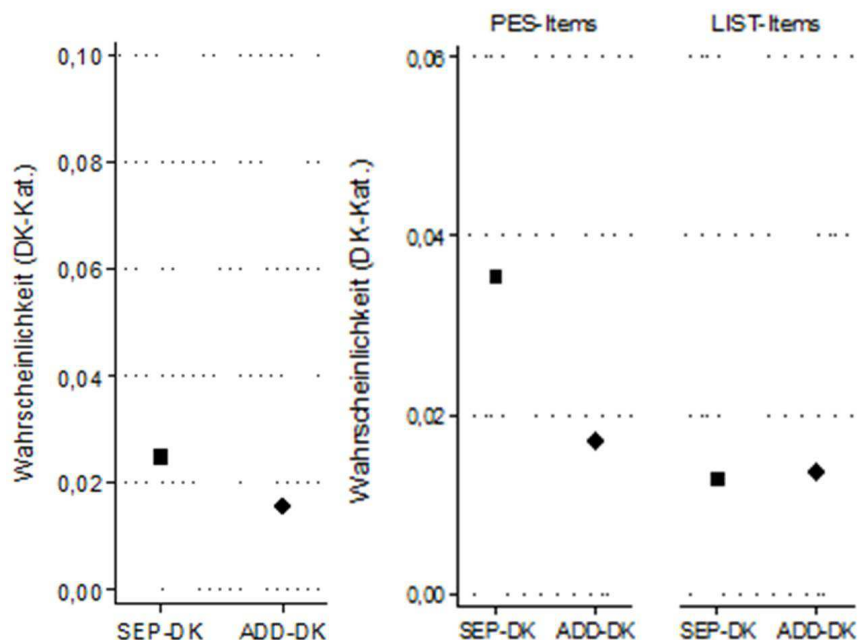


Abbildung 53 Wahl der DK-Kategorie nach Layout (links) und nach Layout und Item-Konstrukt (rechts) in Studie 1

Auch bei der Wahl der rechten Extremkategorie zeigen sich deutlichere Layouteffekte bei den PES-Items als bei den Items über die Lernstrategien (vgl.

Unterabschnitt 9.1.3). Lediglich hinsichtlich der Wahl der Mittelkategorie sind stärkere Effekte des Layouts für die Items über die Lernstrategien zu beobachten (vgl. Unterabschnitt 9.2.3). Generell zeigt sich für diese Items eine sehr hohe Wahrscheinlichkeit der Wahl der mittleren Antwortkategorie, was die Vermutung aufkommen lässt, dass Befragte bei diesen Items, die gegen Ende der Befragung vorgelegt werden, verstärkt ein Satisficing-Verhalten aufweisen. In Studie 2 ist für die Wahl der DK-Kategorie entgegen der Erwartung kein Layouteffekt zu beobachten, obwohl die dort betrachteten Items über die „Einstellungen zur EU“ wie in der PES ebenfalls ein politisches Thema behandeln. Hier könnte die Erfahrung der Panelisten, Einstellungen zu ihnen unvertrauten, abstrakten Themen zu berichten dem Layouteffekt entgegenwirken (Toepoel et al. 2008).

Eine Einschränkung der Analyse der DK-Antworten in den beiden empirischen Studien liegt in den eher geringen Fallzahlen, die zu großen Standardfehlern und breiten Konfidenzintervallen führen und somit ungenaue Schätzungen liefern (Gross und Kriwy 2009). Dadurch ist die Annahme der Nullhypothese („es liegen keine Layouteffekte vor“) mit einer größeren Unsicherheit verbunden. Ein Split-ballot Experiment der variierten graphischen Gestaltung der DK-Vorgabe mit hohen Fallzahlen bzw. mit Items, die mit einer hohen Rate an DK-Antworten beantwortet werden, würden hier tiefergehende Analysen erlauben.

Den Item-Themen „Politische Selbstwirksamkeit“ und „Lernstrategien“ wurde exogen zugewiesen, ob sie als „alltätlich / vertraut“ oder „politisch / abstrakt“ von den Befragten wahrgenommen werden, was für einige Fälle sicherlich nicht zutreffend ist. Deshalb wurde in der Studentenbefragung der Studie 1 auch die individuelle Einstellungsstärke gegenüber den Item-Themen von den Befragten selbst abgefragt. Die individuelle Einstellungsstärke erweist sich über alle drei untersuchten Antwortkategorien als Kovariate von Layouteffekten: Befragte mit einer hohen Einstellungsstärke sind im Vergleich zu Befragten mit einer niedrigen / mittleren Einstellungsstärke in ihrer Antwortabgabe weniger bis hin zu gar nicht empfänglich für das Layout der Ratingskala (vgl. Unterabschnitt 9.1.5, Unterabschnitt 9.2.5 und Unterabschnitt 9.3.5). Diese Interaktion ist stärker bei den Items über die politische Selbstwirksamkeit. Damit unterstützt dieses Ergebnis die Hypothese (H 3), die auf der Annahme basiert, dass mit zunehmender Einstellungsstärke Befragte weniger das Ratingskalenlayout bei dem kognitiven

Fragebeantwortungsprozess einbeziehen und daher mit zunehmender Einstellungsstärke weniger bis hin zu gar keine Layouteffekte mehr auftreten. Zudem sind auch inhaltlich sinnvolle Haupteffekte der Einstellungsstärke auf die Antwortabgabe zu beobachten: Umso stärker die Einstellungsstärke der Befragten ist, desto wahrscheinlicher wählen sie die rechte Extremkategorie (vgl. Unterabschnitt 9.1.4) und desto unwahrscheinlicher die Mittel- bzw. die DK-Kategorie (vgl. Unterabschnitt 9.2.4 bzw. Unterabschnitt 9.3.4) zur Antwortabgabe; dieser Zusammenhang ist unabhängig vom Item-Thema.

Auch bei der Analyse dieser Interaktion der Kovariate „Einstellungsstärke“ und dem Ratingskalenlayout erweisen sich die teilweise geringen Fallzahlen im Bereich der niedrigen Einstellungsstärke als problematisch. Wie bereits zuvor beschrieben führen sie zu ungenauen Schätzern in diesem Bereich. Daher wird für künftige Studien empfohlen, für die Einstellungsstärke einen Index einzusetzen, der eine ausdifferenziertere Erfassung ermöglicht, z.B. durch die Verwendung weiterer Indikatoren wie die Häufigkeit und Ambivalenz.

Unerwartete Ergebnisse zeigen sich in der Online-Befragung der Studie 2 für den Einfluss der Bildung und des Alters auf das Auftreten von Layouteffekten: Entgegen der Hypothese (H 4.2) sind Layouteffekte unabhängig vom Alter der Befragten. Es scheint, dass die abnehmende kognitive Leistungsfähigkeit nicht zu einem verstärkten Rückgriff auf Kontextinformationen und somit auf Elemente des Skalenlayouts führt (vgl. Unterabschnitt 15.3). Für die Bildung zeigt sich wider Erwarten ein zu Hypothese (H 4.1) umgekehrter Zusammenhang: Bei der Wahl der untersuchten Antwortkategorien sind Befragte mit einem hohen Bildungsabschluss empfänglicher für das Layout der Skala und werden davon stärker in ihrer Antwortabgabe beeinflusst als Befragte mit einem niedrigen / mittleren Bildungsabschluss (vgl. Unterabschnitt 15.2). Es scheint, dass gemäß dem Ansatz des Need for Cognition (Cacioppo und Petty 1982) eher Befragte mit einer hohen Bildung aufgrund ihrer umfassenden Informationsverarbeitung alle in der Skala enthaltenen Informationen zur Beantwortung der Frage nutzen, anstatt dass Befragte mit einer niedrigen / mittleren Bildung die nonverbale Information der Skala zur Unterstützung des Informationsverarbeitungsprozesses heranziehen, wie gemäß der Hypothese (H 4.1) erwartet wurde.

Aufgrund geringer Fallzahlen der jungen Befragten und der Befragten ohne und solche mit einem geringen Bildungsabschluss in der Online-Befragung ergeben sich folgende Einschränkungen für die Studie 2: Um eine klare Aussage über die Interaktion zwischen Alter und Layouteffekten treffen zu können, wäre es vorzuziehen, das Alter in Jahren, also als kontinuierliche Variable, in der Untersuchung zu betrachten. Aufgrund geringer Fallzahlen für junge Befragte musste allerdings eine Dichotomisierung der Variable „Alter“ in „unter 55 Jahre“ und in „55 Jahre oder älter“ vorgenommen werden. Auch zum Thema Bildung wäre es interessant zu untersuchen, wie die Gruppe der Befragten ohne Bildungsabschluss durch das Layout beeinflusst wird. Die Fallzahlen für eine differenzierte Betrachtung sich allerdings auch hier nicht groß genug. Des Weiteren sind Bildung und Alter nur zwei Indikatoren der kognitiven Leistungsfähigkeit. Eine ausdifferenziertere Erfassung könnte weitere Erkenntnisse ihrer Interaktion mit der Empfänglichkeit für nonverbale Elemente in Fragebögen bringen.

Eine weitere Einschränkung dieser Dissertation stellen die selektiven Stichproben der beiden empirischen Studien dar. Die Teilnehmer der Studie 1 sind Studenten; es handelt sich also um eine Befragtengruppe, die hinsichtlich ihres Alters und ihrer Bildung homogen sind. Die in dieser Gruppe beobachteten Effekte können nicht auf die Allgemeinbevölkerung übertragen werden. Um einen Schluss über das Auftreten von Layouteffekten in der Allgemeinbevölkerung ziehen zu können, wurde daher in der zweiten empirischen Studie eine Untersuchungsgruppe gewählt, welche die Allgemeinbevölkerung besser repräsentiert. Die Stichprobe der Studie 2 ist repräsentativ für die erwachsene Bevölkerung in Deutschland mit Internetzugang. Diese Stichprobe ist für die jüngeren Altersgruppen nahezu repräsentativ für die Allgemeinbevölkerung in Deutschland, für die älteren Altersgruppen hingegen werden größere Abweichung zu der Allgemeinbevölkerung erwartet: Während in der Gruppe der bis Dreißigjährigen in Deutschland unabhängig von Bildung und sozialer Herkunft nahezu eine vollständige Internetnutzung gegeben ist, hat in der Altersgruppe der über Sechzigjährigen nur noch etwa ein Viertel Zugang zum Internet, hier sind es vor allem jene, mit einer höheren Bildung (Bandilla et al. 2009). Dieser unterschiedliche Bildungs-Bias für die jüngeren und älteren Teilnehmer des GESIS Online Panel Pilots kann die Ergebnisse hinsichtlich des Einflusses des Alters auf die Empfänglichkeit für nonverbale Ratingskalenelemente und somit für Layouteffekte verzerren.

Abschließend wird zusammengefasst, dass die graphische Darstellung von Ratingskalen sowohl in Papierfragebögen als auch in Online-Befragungen die befragten Personen systematisch in ihrem Antwortverhalten beeinflusst. Dieser Effekt beeinflusst nicht nur die Rate des Item Nonresponse und der mittleren Antworten, vielmehr scheint die graphische Darstellung der DK-Kategorie die Wahrnehmung des Charakters der gesamten Ratingskala mit zu beeinflussen. Layouteffekte treten allerdings nicht unter allen Umständen auf. Hauptsächlich wenn es um die Erhebung von Einstellungen zu abstrakteren / unvertrauten Themen geht, greifen Befragte auf den Kontext der Skala und somit auf deren nonverbalen Elemente zurück. Daher kann auch für die Studien von Tourangeau et al. (2004) und Christian et al. (2009) angenommen werden, dass die unterschiedlichen Ergebnisse vorwiegend in den unterschiedlichen Themen (politisches Thema vs. Alltagsfrage) und damit einhergehend in ihrer Salienz begründet sind.

Empfehlungen für zukünftige Forschung

Aus den Ergebnissen dieser Dissertation können folgende Empfehlungen für die künftige Forschung abgeleitet werden:

Das Antwortverhalten wird in dieser Dissertation ausschließlich anhand der Wahl verschiedener Antwortkategorien untersucht. Darüber hinaus bieten Online-Befragungen aufgrund ihrer Para-Daten die Möglichkeit, die durchschnittliche Antwortzeit zu Items als weitere Dimension des Antwortverhaltens zu betrachten. Es ist anzunehmen, dass das Layout den kognitiven Fragebeantwortungsprozess der befragten Personen beeinflusst, was sich in unterschiedlichen Antwortzeiten widerspiegeln könnte. Hier wäre denkbar, dass eine Ratingskala ohne DK-Kategorie, bei der die verbalen und nonverbalen Informationen übereinstimmen, ein schnelleres Beantworten der Items fördern sollte. Im ADD-DK-Layout hingegen bieten die verbalen und nonverbalen Elemente widersprüchliche Informationen, was von den Befragten in einem kognitiv aufwendigeren Prozess verarbeitet werden muss und sie somit in der Fragebeantwortung verlangsamen könnte. Im SEP-DK-Layout wird zwar wie im NO-DK-Layout ebenfalls eindeutige Information an den Befragten über die Skala weitergegeben, allerdings könnte die zusätzlich angebotene DK-Kategorie ebenfalls die Antwortzeit verlängern.

Offen ist die Frage, wie sich der systematische Einfluss des Layouts auf das Antwortverhalten, auf die Güte der Skala der politischen Selbstwirksamkeit und auf die Güte der beiden Subskalen „Zeitmanagement“ und „Aufmerksamkeit“ des LIST-Inventars auswirkt. Hier wäre eine weitere Studie erforderlich, die beispielsweise anhand von konfirmatorischen Faktorenanalysen untersucht, ob das Layout der Antwortformate die Validität und Reliabilität der Skalen, die latente Konstrukte erfassen sollen, beeinflusst.

Um weitere Erkenntnisse darüber zu erlangen, wie Befragte in ihrem Fragebeantwortungsprozess durch das Layout der Ratingskala beeinflusst werden, bieten neben quantitativen auch qualitative Methoden verschiedene Möglichkeiten: Mittels des Verfahrens der kognitiven Interviews können beispielsweise Erkenntnisse darüber gewonnen werden, ob die von Tourangeau et al (2004) entwickelten interpretativen Heuristiken von Befragten als sinnvoll erachtet und in der Frage- bzw. Skaleninterpretation angewandt werden. Viele Teile des kognitiven Fragebeantwortungsprozesses laufen allerdings für den Befragten unbewusst ab, weshalb sich beispielsweise auch das Eye-tracking-Verfahren anbietet, um die Blickbewegungen aufzuzeichnen und anschließend systematisch im Hinblick auf Unterschiede je nach Ratingskalenlayout auszuwerten (für eine Eye-tracking Studie zur Verwendung von Labels in Ratingskalen vgl. Menold et al. 2014).

Die subjektive Einstellungsstärke und die beiden persönlichen Merkmale „Bildung“ und „Alter“ werden in der Dissertation als potentielle Kovariate von Layouteffekten betrachtet. Es wird des Weiteren angenommen, dass ein Satisficing-Verhalten oder eine subjektiv als hoch empfundene Belastung die Empfänglichkeit der Befragten für nonverbale Fragebogen- und Antwortskalenelemente erhöht. Hierzu besteht noch Forschungsbedarf.

Offen ist auch, ob Layouteffekte bei Befragungen auftreten, die von einem Interviewer durchgeführt werden bzw. bei denen ein Interviewer anwesend ist. Zwar dominieren bei dieser Befragungsform die verbalen Elemente, allerdings werden häufig auch in Interviewer administrierten Erhebungen Teile des Erhebungsinstruments, wie beispielsweise Showcards mit Antwortlisten oder Ratingskalen, den Befragten visuell vorgelegt. In manchen Umfrageprogrammen, wie beispielsweise dem International Social Survey Programme (ISSP) Deutschland, ist es zudem üblich, dass der Fragebogen selbständig von den

Befragten am PC ausgefüllt wird (Computer-assisted self-interviewing, CASI), allerdings in Anwesenheit eines Interviewers. Dabei steht dieser für Rückfragen oder bei Problemen mit der Handhabung zur Verfügung. Hier ist die Frage offen, inwieweit Befragte durch das Layout der (zumindest in Teilen) visuell präsentierten Erhebungsinstrumente in Anwesenheit von Interviewern in ihrer Antwortabgabe beeinflusst werden.

Schließlich eröffnet sich noch eine weitere, große Forschungslücke für Layouteffekte in kultur- bzw. ländervergleichenden Studien. Die interpretativen Heuristiken (vgl. Unterabschnitt 1.3) sind im U.S. amerikanischen Kontext entstanden und bisher dort, in den Niederlanden und mit dieser Dissertation auch in Deutschland untersucht worden. Allerdings basieren diese Heuristiken auf der Leserichtung („left and top means first“), auf in der Kultur verankerten Überzeugungen („up means good“) oder auf Erkenntnissen der Wahrnehmungspsychologie (u.a. „near means related“), die selbst wieder dem westlichen Forschungskontext entstammen. Wichtig wäre es, in diesem Zusammenhang zu untersuchen, ob in Kulturen, die sich stark von der westlichen Kultur unterscheiden und / oder die eine andere Schriftsprache verwenden (z.B. Hebräisch, Chinesisch...), die Interpretation der nonverbalen Ratingskalenelemente anders erfolgt. Denn wenn unterschiedliche kulturelle Gruppen durch nonverbale Elemente in unterschiedlicher Weise in ihrer Antwortabgabe beeinflusst werden, kann dies die Vergleichbarkeit der Daten stark einschränken.

Implikationen für die empirische Forschung

Für die Forschungspraxis können folgende Layoutempfehlungen für Ratingskalen in Selbstausfüllern formuliert werden:

1. DK-Kategorie einer Ratingskala abtrennen

Nur bei Verwendung eines Layouts, dessen nonverbale Elemente Informationen bieten, die mit den verbalen oder numerischen Informationen der Ratingskala übereinstimmen, können Fehlinterpretationen der Skala durch die Befragten vermieden werden.

2. Verzicht auf Ausschmückung

Bei der Gestaltung von Selbstausfüller-Fragebögen gibt es keine ausschließlich schmückenden, nonverbalen Elemente; die befragten Personen sprechen allen Elementen im Erhebungsinstrument eine Funktion zu. Dadurch können unerwünschte Effekte entstehen, welche in ihrem Auftreten zudem durch verschiedene Kontextfaktoren modelliert werden. Es wird daher empfohlen, in visuell präsentierten Fragebögen nonverbale Elemente wohlüberlegt einzusetzen. Besonders bei der Gestaltung von Online-Befragungen bietet sich eine reiche Palette an Gestaltungsmöglichkeiten (Farben, Formen, Animationen etc.) an, von der - im Gegensatz zu Papierfragebögen – ohne zusätzliche Kosten Gebrauch gemacht werden kann. Davon ist allerdings abzuraten, da für viele Gestaltungselemente (noch) nicht klar ist, wie sie Befragte beeinflussen.

3. Die Bewertung des Skalenlayouts gehört zur Auswertungsvorbereitung

Jeder Datenanalyse geht eine nähere Betrachtung des Erhebungsinstruments, also der Frage bzw. der Items sowie der Antwortoptionen voraus. Bei diesem Schritt muss auch eine Bewertung des Layouts der Antwortskala, so wie sie dem Befragten tatsächlich bei der Fragebeantwortung vorgelegen hat, vorgenommen werden. Dies ist dringend notwendig, um die späteren Ergebnisse der Datenanalyse richtig bewerten zu können, was insbesondere für vergleichende Studien (international und longitudinal) gilt. Denn die Umsetzung des Erhebungsinstrumentes wird oftmals von Erhebungsinstituten während der Implementation angepasst. Bei der späteren Analyse sollte klar sein, ob zum Beispiel die graphische Darstellung der DK-Kategorie in allen Ländern bzw. über die Zeit identisch ist. Ist dies nicht der Fall, muss der entsprechende Bias in der Auswertung berücksichtigt werden. Daher muss, wie bei allen sozialwissenschaftlichen Daten, die Dokumentation (Metadaten) unbedingt das originale Erhebungsinstrument beinhalten. Denn nur so kann der Effekt des Layout-Bias in der Auswertung berücksichtigt werden.

4. Layout und Instruktionen müssen vollständig sein

Das Weglassen der DK-Kategorie in einer Ratingskala lässt Befragte auf die Mittelkategorie ausweichen. Nur sehr wenige Befragte nutzen die in beiden Studien gegebene Möglichkeit, keine der vorgegebenen Antwortkategorien zu wählen - also eine implizite DK-Antwort zu geben. D.h. Befragte halten sich in der Regel an die

Instruktionen zum Ausfüllen eines Fragebogens (Couper et al. 2004b; Schwarz 1996; Tourangeau et al. 2000). Auch Sturgis et al. (2014) kommen in einer Online-Befragung anhand von nachgeschalteten, offenen Fragen zu dem Ergebnis, dass Befragte ohne Einstellung zu dem Thema bei einer nicht vorgegeben DK-Kategorie die Mittelkategorie wählen, anstatt eine DK-Antwort frei zu berichten. Die Autoren vermuten, dass der Grund für dieses Antwortverhalten das Verschleiern von fehlenden Einstellungen ist. Daher bezeichnen sie diese Mittelkategorie-Antworten auch als „face-saving don’t knows“. Werden in der Analyse diese „face saving don’t knows“ als DK-Antworten gewertet, kommt es zu signifikant unterschiedlichen deskriptiven und multivariaten Ergebnissen als in einer Analyse, in der sie weiterhin als Mittelkategorie-Antworten behandelt werden. Abgeleitet aus den Ergebnissen der Dissertation und der Studie von Sturgis et al. (2014) wird daher empfohlen, bei Einstellungsfragen eine DK-Kategorie anzubieten. Andernfalls muss damit gerechnet werden, dass Befragte systematisch auf andere, inhaltliche Antwortkategorien ausweichen, wie eben die Mittelkategorie, was schließlich zu einer Verzerrung der Ergebnisse führt.²⁷

²⁷ Für eine ausführliche Diskussion um die Verwendung von DK-Kategorien in Ratingskalen siehe auch (Menold und Bogner 2012,2014).

LITERATURVERZEICHNIS

- Alwin, Duane F., und Jon A. Krosnick*, 1991: The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods and Research* 20: S. 139-181.
- Amoo, T., und H. H. Friedman*, 2001: Do numeric values influence subjects' responses to rating scales? *Journal of International Marketing and Marketing Research* 26: S. 41-46.
- Andrews, Frank M.*, 1984: Construct validity and error components of survey measures: a structural modeling approach. *Public Opinion Quarterly* 48: S. 409-442.
- Armitage, C., und C. Deeptose*, 2004: Changing student evaluations by means of the numeric values of rating scales. *Psychological Learning and Teaching* 3: S. 122-125.
- Aronson, Elliot, Timothy Wilson und Robin M. Akert*, 2008: *Sozialpsychologie*. München: Pearson Studium.
- Asendorpf, J. B., und F. J. Neyer*, 2012: *Psychologie der Persönlichkeit*. Heidelberg: Springer.
- Atkeson, Lonna Rae, Alex N. Adams und R. Michael Alvarez*, 2014: Nonresponse and mode effects in self- and interviewer-administered survey. *Political Analysis* 22: S. 304-320.
- Auspurg, Katrin, und Thomas Hinz*, 2010: Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen - Probleme und Fehleinschätzungen am Beispiel von Bildungschancen im Kohortenverlauf. *Zeitschrift für Soziologie* 40: S. 62-73.
- Backhaus, Klaus, Bernd Erichson, Wulff Plinke und Rolf Weiber*, 2006: *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Bandilla, Wolfgang, Lars Kaczmarek, Michael Blohm und Wolfgang Neuert*, 2009: Coverage- und Nonresponse-Effekte bei Online-Bevölkerungsumfragen. S. 129-144 in: *Jackob, Nikolaus, Harald Schoen und Thomas Zerback* (Hg.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bassili, John N., und Jon A. Krosnick*, 2000: Do strength-related attitude properties determine susceptibility to response effects? New evidence from response latency, attitude extremity, and aggregate indices. *Political Psychology* 21: S. 107-132.
- Best, Henning, und Christof Wolf*, 2010: Logistische Regression. S. 829-854 in: *Wolf, Christof, und Henning Best* (Hg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag.

- Bishop, George, und Andrew Smith*, 2001: Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly* 65: S. 479-505.
- Bishop, George F.*, 1990: Issue involvement and response effects in public opinion surveys. *Public Opinion Quarterly* 54: S. 209-218.
- Bishop, George F., Robert W. Oldendick und Alfred J. Tuchfarber*, 1983: Effects of filter questions in public opinion surveys. *Public Opinion Quarterly* 47: S.
- Bless, Herbert, Fritz Strack und Norbert Schwarz*, 1993: The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology* 23: S.
- Bogner, Kathrin, und Uta Landrock*, 2014: Antworttendenzen in standardisierten Umfragen. *SDM Survey Guidelines*. Gesis. Mannheim.
- Bowling, Ann*, 2005: Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health* 27: S. 281-291.
- Bradburn, N. M., und Seymour Sudman*, 1988: Polls and surveys: Understanding what they tell us San Fransisco: Jossey-Bass.
- Cacioppo, John T., und Richard Petty, E.*, 1982: The need for cognition. *Journal of Personality and Social Psychology* 42: S. 116-131.
- Cameron, A. C., und P. K. Trivedi*, 1998: Regression analysis of count data. Cambridge: Cambridge University Press.
- Christian, Leah Melani, und Don A. Dillman*, 2004: The influence of symbolic and graphical language manipulations on answers to paper self-administered questionnaires. *Public Opinion Quarterly* 68: S. 57-80.
- Christian, Leah Melani, Don A. Dillman und Jolene D. Smyth*, 2007: Helping respondents get it right the first time: The influence of words, symbols, and graphics in web surveys. *Public Opinion Quarterly* 71: S. 113-125.
- Christian, Leah Melani, Don A. Dillman und Jolene D. Smyth*, 2008: The effects of mode and format on answers to scalar questions in telephone and web surveys. S. 250-275 in: *Lepkowski, J., C. Tucker, M. Brick, Edith D. De Leeuw, L. Japec, Paul Lavrakas, M. Link und Roberta L. Sangster* (Hg.), *Advances in Telephone Survey Methodology*. New York: Wiley-Interscience.
- Christian, Leah Melani, Nicholas L. Parsons und Don A. Dillman*, 2009: Designing scalar questions for web surveys. *Sociological Methods & Research* 37: S. 393-425.
- Converse, J. M., und Stanley Presser*, 1986: Survey questions. Beverly Hills: Sage Publications, Inc.
- Converse, Philip E.*, 2006: The nature of belief systems in mass publics (1964). *Critical Review: A Journal of Politics and Society* 18: S. 1-74.
- Coombs, Clyde H., und Lolagene C. Coombs*, 1976: "Don't know": Item ambiguity or respondents uncertainty? *Public Opinion Quarterly* 40: S. 497-514.
- Couper, Mick P., Michael W. Traugott und Mark Lamias, J.*, 2001: Web survey design and administration. *Public Opinion Quarterly* 65: S. 230-253.

- Couper, Mick P., Roger Tourangeau und Kristin Kenyon*, 2004a: Picture this! Exploring visual effects in web surveys. *Public Opinion Quarterly* 68: S. 255-266.
- Couper, Mick P., Roger Tourangeau, Frederick G. Conrad und S. D. Crawford*, 2004b: What they see is what we get - Response options for web surveys. *Social Science Computer Review* 22: S. 111.
- Craig, Stephen C., Richard G. Niemi und Glenn E. Silver*, 1990: Political efficacy and trust: A report on the NES pilot study items. *Political Behavior* 12: S. 289-314.
- Davies, Bethan L.*, 2000: Grice's cooperative principle: meaning and rationality. *Journal of Pragmatics* 39: S. 2308-2331.
- De Leeuw, Edith D.*, 1992: Data quality in mail, telephone, and face-to-face surveys. Amsterdam: TT-Publicaties.
- De Leeuw, Edith D.*, 2005: To mix or not to mix data collection methods in surveys. *Journal of Official Statistics*. 21: 233-255.
- De Leeuw, Edith D., Don A. Dillman und Joop J. Hox*, 2008: Mixed mode surveys: When and why. S. 299-316 in: *De Leeuw, Edith D., Joop J. Hox und Don A. Dillman* (Hg.), *International Handbook of Survey Methodology*. New York: Taylor & Francis Group.
- Deary, Ian J., Steve Strand, Pauline Smith und Cres Fernandes*, 2007: Intelligence and educational achievement. *Intelligence* 35: S. 13-21.
- Derouvray, Cristel, und Mick P. Couper*, 2002: Designing a strategy for reducing "no opinion" responses in web-based surveys. *Social Science Computer Review* 20: S. 3-9.
- Dillman, Don A., und Leah Melani Christian*, 2005: Survey mode as a source of instability in responses across surveys. *Field Methods* 17: S. 30-52.
- Dillman, Don A., Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, Jodi Berck und Benjamin L. Messer*, 2009: Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research* 38: S. 1-18.
- Eagly, Alice Hendrickson, und Shelly Chaiken*, 1993: *The psychology of attitudes*. Fort Worth: Harcourt, Brace, & Janovich.
- Eliason, Scott R.*, 1993: *Maximum likelihood estimation: Logic and practice*. Los Angeles: Sage Publications, Inc.
- Faulkenberry, G. David Mason Robert*, 1978: Characteristics of nonopinion and no opinion response groups. *Public Opinion Quarterly* 42: S.
- Feick, Lawrence F.*, 1989: Latent class analysis of survey questions that include don't know responses. *Public Opinion Quarterly* 53: S. 525-547.
- Ferber, Robert*, 1966: Item nonreponse in a consumer survey. *Public Opinion Quarterly* 30: S. 399-415.
- Fowler, F. J., und C. F. Cannell*, 1996: Using behavioral coding to identify cognitive problems with survey questions. S. 15-36 in: *Schwarz, Norbert, und Seymour Sudman* (Hg.), *Answering questions: methodology for determining*

- cognitive and communicative processes in survey research. San Francisco: Jossey-Bass.
- Franzen, Axel*, 2014: Antwortskalen in standardisierten Umfragen. S. 701-712 in: *Bauer, Nina , und Jörg Blasius* (Hg.), Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS.
- Fricker, Scott, Mirta Galesic, Roger Tourangeau und Ting Yan*, 2005: An experimental comparison of web and telephon surveys. *Public Opinion Quarterly* 69: S. 370-392.
- Friedman, L. W., und H. H. Friedman*, 1994: A comparison of vertical and horizontal rating scales. *The Mid-Atlantic Journal of Business* 30: S. 107-202.
- Gabler, Siegfried, und Sabine Häder*, 2002: Idiosyncrasies in telephone sampling - the case of Germany. *International Journal of Public Opinion Research* 14: S. 339-345.
- Gabler, Siegfried, Sabine Häder, Iris Lehnhoff und Elisabeth Mardian*, 2012: Weighting for unequal inclusion probabilities and nonresponse in dual frame telephone surveys. S. in: *Häder, Sabine, Michael Häder und Mike Kühne* (Hg.), Telephone surveys in Europe: Research and practice. Heidelberg: Springer.
- Galesic, Mirta, Roger Tourangeau, Mick P. Couper und Frederick G. Conrad*, 2008a: Eye-tracking data. New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly* 72: S. 892-913.
- Galesic, Mirta, Roger Tourangeau, Mick P. Couper und Frederick G. Conrad*, 2008b: Eye tracking data. New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly* 72: S. 892-913.
- Gautschi, Thomas*, 2010: Maximum-Likelihood Schätztheorie. S. in: *Best, Henning , und Christof Wolf* (Hg.), Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gilljam, Mikeal, und Donald Granberg*, 1993: Should we take don't know for an answer! *Public Opinion Quarterly* 57: S. 348-357.
- Grice, Paul*, 1993: Logik und Konversation. S. 243-265 in: *Meggle, Georg* (Hg.), Handlung, Kommunikation, Bedeutung. Frankfurt am Main: Suhrkamp.
- Gross, Christiane, und Peter Kriwy*, 2009: Kleine Fallzahlen in der empirischen Sozialforschung. S. 9-21 in: *Kriwy, Peter , und Christiane Gross* (Hg.), Quantitative empirische Sozialforschung mit kleinen Fallzahlen. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Groves, Robert M., und Robert Louis Kahn*, 1979: Surveys by telephone: A national comparison with personal interviews. New York: Academic Press.
- Hartley, James, und Lucy R. Betts*, 2010: Four layouts and a finding: the effects of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology* 13: S. 17-27.

- Heerwegh, Dirk*, 2009: Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research* 21: S. 111-121.
- Heerwegh, Dirk, und Geert Loosveldt*, 2008: Face-to-face versus web surveying in a high-internet-coverage population. Differences in response quality. *Public Opinion Quarterly* 72: S. 836-846.
- Hippler, Hans-Jürgen, und Norbert Schwarz*, 1989: 'No opinion'-filters: A cognitive perspective. *International Journal of Public Opinion Research* 1: S. 77-87.
- Holbrook, Allyson L., Melanie C. Green und Jon A. Krosnick*, 2003: Telephon versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly* 67: S. 79-125.
- Hosmer, David W., und Stanley Lemeshow*, 2000: Applied logistic regression. Hoboken: John Wiley & Sons, Inc.
- Hox, Joop J.*, 2010: Multilevel analysis: Techniques and applications. New York: Routledge.
- Huber, Peter J.*, 1967: The bahavior of maximum likelihood estimates under nonstandard conditions. Fifth Berkeley Symposium on Mathematical Statistics and Probability.
- Jenkins, Cleo R., und Don A. Dillman*, 1995: Towards a theory of self-administered questionnaire design. S. in: *Lyberg, Lars; Biemer, Paul; Collins, M.; Deleeuw, Edith; Dippo, C.; Schwarz, Norbert; Trewin, D.* (Hg.), *Survey Measurement and Process Quality*. New York: Wiley-Interscience.
- Johanson, George A., Crystal J. Gips und Charles E. Rich*, 1993: "If you can't say something nice": A variation on the social desirability response set. *Evaluation Review* 17: S. 116-122.
- Katz, Daniel*, 1942: Do interviewers bias pool results? *Public Opinion Quarterly* 6: S. 248-268.
- Kerlinger, Fred N.*, 1973: Foundations of behavioral research. New York: Holt, Rinehart and Winston, Inc.
- Klausch, Thomas, Joop J. Hox und Barry Schouten*, 2013: Measurement effects of survey mode on the equivalence of attitudinal rating scale questions. *Sociological Methods and Research* 42: S. 227-263.
- Knäuper, Bärbel*, 1999: The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly*: S. 347-370.
- Koffka, Kurt*, 1935: Principles of Gestalt psychology. London: Lund Humphries.
- Krosnick, Jon A.*, 1991: Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5: S. 213-236.
- Krosnick, Jon A.*, 1992: The impact of cognitive sophistication and attitude importance on response order effects and question order effects. S. 203-218 in: *Schwarz, Norbert , und Seymour Sudman* (Hg.), *Order effects in social and psychological research*. New York: Springer-Verlag.

- Krosnick, Jon A.*, 1999: Survey research. *Annual Review of Psychology* 50: S. 537-567.
- Krosnick, Jon A.*, 2002: The causes of no-opinion responses to attitude measures in surveys: they are rarely what they appear to be. S. 88-100 in: *Groves, Robert M., Don A. Dillman, John N. Eltinge und Roderick J. A. Little* (Hg.), *Survey nonresponse*. New York: Wiley-Interscience.
- Krosnick, Jon A., und Duane F. Alwin*, 1987: An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly* 51: S. 201-219.
- Krosnick, Jon A., und Robert P. Abelson*, 1991: The case for measuring attitude strength. S. 177-203 in: *Tanur, Judith M.* (Hg.), *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russel Sage Foundation.
- Krosnick, Jon A., und Wendy R. Smith*, 1994: Attitude strength. *The Encyclopedia of Human Behavior*. Ramachandran, V. S. San Diego. Academic Press.
- Krosnick, Jon A., und Richard Petty, E.*, 1995: Attitude strength: an overview. S. in: *Petty, Richard, E. , und Jon A. Krosnick* (Hg.), *Attitude strength: antecedents and consequences*. New York: Psychology Press.
- Krosnick, Jon A., und Leandre R. Fabrigar*, 1997: Designing Rating Scales for Effective Measurement in Surveys. S. 141-164 in: *Lyberg, Lars, Paul Biemer, Martin Collins, Edith Leeuw, Cathryn Dippo, Norbert Schwarz und Dennis Trewin* (Hg.), *Survey Measurement and Process Quality*. New York: Wiley-Interscience.
- Krosnick, Jon A., und Stanley Presser*, 2010: Question and Questionnaire Design. S. 263-313 in: *Marsden, Peter V. , und James D. Wright* (Hg.), *Handbook of Survey Research 2*. Bingley: Emerald Group Publishing.
- Krosnick, Jon A., S. S. Narayan und W. R. Smith*, 1996: Satisficing in surveys: initial evidence. S. 29-44 in: *Braverman, M. T. , und J. K. Slater* (Hg.), *Advances in survey research*. San Francisco: Jossey-Bass.
- Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, Hanemann Michael W., Raymond J. Kopp, Robert C. Mitchel, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green und Michael Conaway*, 2002: The impact of "no opinion" response options on data quality. Non-attitude reduction or ann invitation to satisfice? *Public Opinion Quarterly* 66: S. 371-402.
- Kulas, John T., Alicia A. Stachowski und Brad A. Haynes*, 2008: Middle response functioning in likert-responses to personality items. *Journal of Business and Psychology* 22: S. 251-259.
- Kwak, Nojin, und Barry Radler*, 2002: A comparison between mail and web surveys: response pattern, respondent profile, and data quality. *Journal of Offical Statistics* 18: S. 257-273.
- Lavine, Howard, Joseph W. Huff und Stephen H. Wagner*, 1998: The moderating influence of attitude strength on the susceptibility to context effects in attitude surveys. *Journal of Personality and Social Psychology* 75: S. 359-373.

- Likert, R.*, 1932: A technique for the measurement of attitudes. *Archives of Psychology* 140: S. 1-55.
- Long, Scott J., und Jeremy Freese*, 2006: Regression models for categorical dependent variables using Stata. College Station, Texas: Stata Press.
- Mcclendon, Mckee J., und Duane F. Alwin*, 1993: No-opinion filters and attitude measurement reliability. *Sociological Methods and Research* 21: S. 438-464.
- Menold, Natalja*, 2014: Does the polarity of rating scales matter? How unipolar, bipolar and mixed rating scales affect measurements with latent variables. 18th ISA World Congress of Sociology. Yokohama, Japan.
- Menold, Natalja, und Kathrin Bogner*, 2012: Antwortskalen in sozialwissenschaftlichen Umfragen: Theoretische Modelle, Stand der Forschung und Forschungsperspektiven. S. in: *Soeffner, Hans-Georg* (Hg.), *Transnationale Vergesellschaftungen: Verhandlungen des 35. Kongresses der Deutschen Gesellschaft für Soziologie in Frankfurt am Main 2010*. Wiesbaden: Springer VS.
- Menold, Natalja, und Kathrin Bogner*, 2014: Gestaltung von Ratingskalen in Fragebögen. *SDM Survey Guidelines*. Gesis. Mannheim.
- Menold, Natalja, Lars Kaczmirek, Timo Lenzner und Ales Neusar*, 2014: How do respondents attend to verbal labels in rating scales? *Field Methods* 26: S. 21-39.
- Payne, S. L.*, 1950: Thoughts about meaningless questions. *Public Opinion Quarterly* 14: S. 687-696.
- Poe, Gail S., Isadore Seeman, Joseph McLaughlin, Eric Mehl und Michael Dietz*, 1988: 'Don't know' boxes in factual questions in a mail questionnaire: Effects on level and quality of response. *Public Opinion Quarterly* 52: S.
- Porst, Rolf*, 2009: Fragebogen. Ein Arbeitsbuch. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rabe-Hesketh, Sophia, und Anders Skrondal*, 2008: Multilevel and longitudinal modeling using stata. College Station: Stata Press.
- Rapoport, Ronald B.*, 1982: Sex differences in attitude expression: A generational explanation. *Public Opinion Quarterly* 46: S. 86-96.
- Rapoport, Ronald B.*, 1985: Like mother, like daughter: Intergenerational transmission of DK response rates. *Public Opinion Quarterly* 49: S. 198-208.
- Redline, Cleo, und Don A. Dillman*, 2002: The influence of alternative visual designs on respondents' performances with branching instructions in self-administered questionnaires. S. in: *Groves, Robert, Don A. Dillman, J. Eltinge und R. Little* (Hg.), *Survey Nonresponse*. New York: John Wiley and Sons, Inc.
- Revilla, Melanie A., und Willem E. Saris*, 2013: A comparison of the quality of questions in a face-to-face and a web survey. *International Journal of Public Opinion Research* 25: S. 242-253.

- Rockwood, Todd H., Roberta L. Sangster und Don A. Dillman*, 1997: The effect of response categories on questionnaire answers: context and mode effects. *Sociological Methods and Research* 26: S. 118-140.
- Salthouse, Timothy A.*, 1991: Theoretical perspectives on cognitive aging. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Salthouse, Timothy A.*, 2004: What and when of cognitive aging. *Current Directions in Psychological Science* 13: S. 140-144.
- Sanchez, Maria Elena, und Giovanna Morchio*, 1992: Probing don't know answers. Effects on survey estimates and variable relationships. *Public Opinion Quarterly* 56: S. 454-474.
- Scherpenzeel, A. C., und Willem E. Saris*, 1997: The validity and reliability of survey questions: a meta-analysis of MTMM studies. *Sociological Methods and Research* 25: S. 341-383.
- Schnell, Rainer, Paul B. Hill und Elke Esser*, 2011: Methoden der empirischen Sozialforschung. München: Oldenbourg Verlag.
- Schonlau, M., A. Van Soest, A. Kapteyn und Mick P. Couper*, 2009: Selection bias in web surveys and the use of propensity scores. *Sociological Methods and Research* 37: S. 291-318.
- Schuman, Howard, und Stanley Presser*, 1981: Questions and answers in attitude surveys: experiments on question form, wording, and context. New York: Springer-Verlag.
- Schwarz, Norbert*, 1996: Cognition and communication: Judgmental biases, research methods and the logic of conversation. Hillsdale, New York: Erlbaum.
- Schwarz, Norbert, Bärbel Knäuper, Hans-Jürgen Hippler, Elisabeth Noelle-Neumann und Leslie Clark*, 1991: Rating Scales-Numeric values may change the meaning of scale labels. *Public Opinion Quarterly* 55: S. 570-582.
- Shih, Tse-Hua, und Xitao Fan*, 2008: Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods* 20: S. 249-271.
- Smith, Tom W.*, 1995: Little things matter: A sampler of how differences in questionnaire format can affect survey responses. *Proceedings of the American Statistical Association, Survey Research Methods Section*: 1046-1051.
- Smyth, Jolene D., Don A. Dillman und Leah Melani Christian*, 2006: Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly* 70: S. 66-77.
- Snijders, Tom A. B., und Roel J. Bosker*, 2012: Multilevel analysis. An introduction to basic and advanced multilevel modeling. London: Sage Publications Inc.
- Stern, Michael J., Don A. Dillman und Jolene D. Smyth*, 2007: Visual design, order effects, and respondent characteristics in a self-administered survey. *Survey Research and Methods* 1: S. 121-138.
- Struminskaya, Bella, Lars Kaczmirek, Ines Schaurer und Wolfgang Bandilla*, 2014: Assessing representativeness of a probability-based online panel in Germany. S. 61-84 in: *Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja*

- Göritz, Jon A. Krosnick und Paul J. Lavrakas (Hg.), Online Panel Research: A Data Quality Perspective. New York: Wiley Series in Survey Methodology.
- Sturgis, Patrick, Caroline Roberts und Patten Smith, 2014: Middle response alternatives revisited: how the neither/nor response acts as a way of saying "I don't know"? Sociological Methods & Research 43: S. 15-38.
- Thurstone, L. L., und E. J. Chave, 1929: The measurement of attitude. Chicago: University of Chicago Press.
- Toepoel, Vera, und Don A. Dillman, 2011: Words, numbers, and visual heuristics in web surveys: Is there a hierarchy of importance? Social Science Computer Review 29: S. 193-207.
- Toepoel, Vera, Marcel Das und Van Arthur Soest, 2006: Design of web questionnaires: the effect of layout in rating scales. CentER Discussion Paper. Tilburg University.
- Toepoel, Vera, Marcel Das und Van Arthur Soest, 2008: Effects of design in web surveys. Comparing trained and fresh respondents. Public Opinion Quarterly 72: S. 985-1007.
- Toepoel, Vera, Marcel Das und Van Arthur Soest, 2009: Design of web questionnaires: The effect of layout in rating scales. Journal of Official Statistics 25: S. 509-528.
- Tourangeau, Roger, und Ting Yan, 2007: Sensitive questions in surveys. Psychological Bulletin 133: S. 859-883.
- Tourangeau, Roger, Lance J. Rips und Kenneth Rasinski, 2000: The psychology of survey response. Cambridge: Cambridge University Press.
- Tourangeau, Roger, Mick P. Couper und Frederick G. Conrad, 2004: Spacing, position, and order. Interpretative heuristics for visual features of survey questions. Public Opinion Quarterly 68: S. 368-393.
- Tourangeau, Roger, Mick P. Couper und Frederick G. Conrad, 2007: Color, labels, and interpretative heuristics for response scales. Public Opinion Quarterly 71: S. 91-112.
- Trautwein, Ulrich, Herbert W. Marsh, Benjamin Nagengast, Oliver Lüdtke, Gabriel Nagy und Kathrin Jonkmann, 2012: Probing for the multiplicative term in modern expectancy - value theory: A latent interaction modeling study. Journal of Educational Psychology 104: S. 763-777.
- Tutz, Gerhard, 2010: Regressionen für Zählvariablen. S. 887-904 in: Wolf, Christof, und Henning Best (Hg.), Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Vaillancourt, P. M., 1973: Stability of children's survey responses. Public Opinion Quarterly 37: S. 373-387.
- Visser, Penny, Jon A. Krosnick, Jesse Marquette und Michael Curtin, 2000: Improving election forecasting: Allocation of undecided respondents, identification of likely voters, and response order effects. S. in: Lavrakas, Paul, und Michael W. Traugott (Hg.), Election polls, the news media, and democracy. New York: Chatham House.

- Vroom, V. H.*, 1964: Work and Motivation. New York: Wiley.
- Wertheimer, Max*, 1923: Untersuchungen zur Lehre von der Gestalt Psychologische Forschung 4: S. 301-350.
- White, Halbert*, 1980: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: S. 817-838.
- Wild, K.-P., und U. Schiefele*, 1994: Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie* 15: S. 185-200.
- Wooldridge, Jeffrey M.*, 2002: Econometric analyses of cross section and panel data. Cambridge: The MIT Press.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, die vorliegende Dissertation selbständig angefertigt und mich keiner anderen als der in ihr angegebenen Hilfsmittel bedient zu haben. Insbesondere sind sämtliche Zitate aus anderen Quellen als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, den 09. Dezember 2014

Kathrin Bogner, geb. Wohn

ANHANG

Tabelle A 1 Verteilung des Item 1 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	0,9 %	1,0 %	5,2 %
	lehne eher ab	35,5 %	27,2 %	27,8 %
	weder noch	15,9 %	11,7 %	15,5 %
	stimme eher zu	38,3 %	41,7 %	33,0 %
	stimme stark zu	9,3 %	14,6 %	15,5 %
	kann ich nicht sagen	-	3,9 %	2,1 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-1,31 (0,46)	-1,18 (0,48)	-1,14 (0,49)
S (SE)		0,03 (0,23)	-0,26 (0,25)	-0,15 (0,25)
Mean (SD)		3,20 (1,06)	3,43 (1,09)	3,27 (1,19)
Shapiro-Wilk		0,842***	0,858***	0,849***
N		107	99	94

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 2 Verteilung des Item 2 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	5,6 %	8,7 %	10,3 %
	lehne eher ab	32,7 %	23,3 %	19,6 %
	weder noch	14,0 %	14,6 %	11,3 %
	stimme eher zu	34,6 %	35,0 %	26,8 %
	stimme stark zu	13,1 %	12,6 %	26,8 %
	kann ich nicht sagen	-	5,8%	4,1 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-1,20 (0,46)	-1,06 (0,49)	-1,18 (0,50)
S (SE)		-0,06 (0,23)	-0,27 (0,25)	-0,39 (0,25)
Mean (SD)		3,17 (1,19)	3,21 (1,22)	3,42 (1,38)
Shapiro-Wilk		0,879***	0,884***	0,860***
N		107	97	92

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 3 Verteilung des Item 3 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	26,2 %	21,4 %	23,7 %
	lehne eher ab	42,1 %	49,5 %	40,2 %
	weder noch	17,8 %	14,6 %	18,6 %
	stimme eher zu	14,0 %	11,7 %	13,4 %
	stimme stark zu	-	1,9 %	2,1 %
	kann ich nicht sagen	-	1,0 %	1,0 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-0,71 (,46)	0,19 (0,47)	-0,39 (0,49)
S (SE)		0,50 (,23)	0,80 (0,24)	0,59 (0,25)
Mean (SD)		2,20 (,99)	2,23 (0,98)	2,28 (1,05)
Shapiro-Wilk		0,848***	0,840***	0,864***
N		107	102	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 4 Verteilung des Item 4 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	8,4 %	7,8 %	8,2 %
	lehne eher ab	22,4 %	32,0 %	35,1 %
	weder noch	26,2 %	20,4 %	14,4 %
	stimme eher zu	42,1 %	33,0 %	36,1 %
	stimme stark zu	0,9 %	3,9 %	4,1 %
	kann ich nicht sagen	-	2,9 %	1,0 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-0,83 (0,46)	-1,06 (0,48)	-1,21 (0,49)
S (SE)		-0,54 (0,23)	-0,06 (0,24)	-0,04 (0,25)
Mean (SD)		3,05 (1,01)	2,93 (1,08)	2,93 (1,11)
Shapiro-Wilk		0,840***	0,881 ***	0,862 ***
N		107	100	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 5 Verteilung des Item 5 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	31,8 %	29,1 %	36,1 %
	lehne eher ab	34,6 %	38,8 %	35,1 %
	weder noch	13,1 %	14,6 %	7,2 %
	stimme eher zu	15,0 %	13,6 %	14,4 %
	stimme stark zu	4,7 %	2,9 %	5,2 %
	kann ich nicht sagen	-	1,0 %	1,0 %
	keine Markierung	0,9 %	-	1,0 %
N		107	103	97
K (SE)		-0,52 (0,47)	-0,32 (0,47)	-0,29 (0,49)
S (SE)		0,72 (0,24)	0,73 (0,24)	0,90 (0,25)
Mean (SD)		2,25 (1,20)	2,22 (1,11)	2,16 (1,22)
Shapiro-Wilk		0,850***	0,853***	0,813***
N		106	102	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 6 Verteilung des Item 6 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	14,0 %	8,7 %	15,5 %
	lehne eher ab	34,6 %	34,0 %	26,8 %
	weder noch	23,4 %	17,5 %	16,5 %
	stimme eher zu	23,4 %	22,3 %	27,8 %
	stimme stark zu	4,7 %	5,8 %	6,2 %
	kann ich nicht sagen	-	10,7 %	5,2 %
	keine Markierung	-	1,0 %	1,0 %
	doppelte Nennung	-	-	1,0 %
N		107	103	97
K (SE)		-0,88 (0,46)	-0,92 (0,50)	-1,17 (0,50)
S (SE)		0,20 (0,23)	0,26 (0,25)	0,03 (0,25)
Mean (SD)		2,70 (1,12)	2,80 (1,13)	2,81 (1,23)
Shapiro-Wilk		0,902***	0,885 ***	0,896 ***
N		107	91	90

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 7 Verteilung des Item 7 der PES

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	3,7 %	-	4,1 %
	lehne eher ab	8,4 %	9,7 %	12,4 %
	weder noch	22,4 %	15,5 %	18,6 %
	stimme eher zu	47,7 %	53,4 %	34,0 %
	stimme stark zu	17,8 %	14,6 %	27,8 %
	kann ich nicht sagen	-	6,8 %	2,1 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		0,46 (0,46)	0,11 (0,49)	-0,41 (0,49)
S (SE)		-0,79 (0,23)	-0,67 (0,25)	-0,65 (0,25)
Mean (SD)		3,67 (0,99)	3,78 (0,84)	3,71 (1,14)
Shapiro-Wilk		0,862***	0,824***	0,873***
N		107	96	94

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 8 Deskriptive Statistik des summierten Index der PES-Items

	NO-DK	Layout SEP-DK	ADD-DK
K (SE)	0,51 (0,47)	-0,44 (0,54)	-0,30 (0,53)
S (SE)	-0,11 (0,24)	-0,12 (0,27)	-0,14 (0,27)
Mean (SD)	2,89 (0,37)	2,94 (0,33)	2,94 (0,34)
Shapiro-Wilk	0,971**	0,972*	0,976
N	106	78	82

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 9 Ergebnisse der Varianzanalyse der PES-Items

	NO-DK	Layout SEP-DK	ADD-DK	
	Mean (SD)	Mean (SD)	Mean (SD)	F-Test
Item 1	3,20 (1,06)	3,43 (1,09)	3,27 (1,12)	$F_{(2, 297)} = 1,23$ $p = 0,294$
Item 2	3,17 (1,19)	3,21 (1,22)	3,42 (1,38)	$F_{(2, 293)} = 1,15$ $p = 0,318$
Item 3	2,2 (0,99)	2,23 (0,98)	2,28 (1,05)	$F_{(2, 301)} = 0,20$ $p = 0,821$
Item 4	3,05 (1,01)	2,93 (1,08)	2,93 (1,11)	$F_{(2, 299)} = 0,43$ $p = 0,653$
Item 5	2,25 (1,20)	2,22 (1,11)	2,16 (1,22)	$F_{(2, 300)} = 0,17$ $p = 0,843$
Item 6	2,70 (1,12)	2,80 (1,13)	2,81 (1,23)	$F_{(2, 285)} = 0,28$ $p = 0,754$
Item 7	3,67 (0,99)	3,78 (0,84)	3,71 (1,14)	$F_{(2, 294)} = 0,30$ $p = 0,738$
Summierter PES-Index	2,89 (0,37)	2,94 (0,33)	2,94 (0,34)	$F_{(2, 263)} = 0,51$ $p = 0,601$

Tabelle A 10 Verteilung des Item 1 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	16,8 %	10,7 %	18,6 %
	trifft wenig zu	26,2 %	27,2 %	28,9 %
	trifft teilweise zu	30,8 %	35,0 %	25,8 %
	trifft ziemlich zu	21,5 %	19,4 %	19,6 %
	trifft voll und ganz zu	4,7 %	5,8 %	5,2 %
	kann ich nicht sagen	-	1,9 %	1,0 %
	keine Markierung	-	-	1,0 %
	N	107	103	97
	K (SE)	-0,82 (0,46)	-0,53 (0,47)	-0,85 (0,49)
	S (SE)	0,7 (0,23)	0,11 (0,24)	0,22 (0,25)
	Mean (SD)	2,71 (1,12)	2,82 (1,06)	2,63 (1,16)
	Shapiro-Wilk	0,908***	0,913***	0,908***
	N	107	101	95

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Tabelle A 11 Verteilung des Item 2 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	8,4 %	13,6 %	15,5 %
	trifft wenig zu	25,2 %	27,2 %	20,6 %
	trifft teilweise zu	25,2 %	15,5 %	20,6 %
	trifft ziemlich zu	29,9 %	35,0 %	28,9 %
	trifft voll und ganz zu	11,2 %	7,8 %	13,4 %
	kann ich nicht sagen	-	1,0 %	-
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-0,90 (0,46)	-1,18 (0,47)	-1,12 (0,49)
S (SE)		-0,9 (0,23)	-0,12 (0,24)	-0,14 (0,25)
Mean (SD)		3,10 (1,16)	2,96 (1,23)	3,04 (1,16)
Shapiro-Wilk		0,911***	0,887***	0,902***
N		107	102	96

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 12 Verteilung des Item 3 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	28,0 %	27,2 %	33,0 %
	trifft wenig zu	41,1 %	32,0 %	35,1 %
	trifft teilweise zu	18,7 %	28,2 %	16,5 %
	trifft ziemlich zu	8,4 %	9,7 %	8,2 %
	trifft voll und ganz zu	3,7 %	1,9 %	5,2 %
	kann ich nicht sagen	-	1,0 %	1,0 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		0,25 (0,46)	-0,48 (0,48)	0,13 (0,49)
S (SE)		0,84 (0,23)	0,44 (0,24)	0,91 (0,25)
Mean (SD)		2,19 (1,06)	2,26 (1,03)	2,16 (1,14)
Shapiro-Wilk		0,854***	0,881***	0,842***
N		107	102	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 13 Verteilung des Item 4 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	20,6 %	24,3 %	25,8 %
	trifft wenig zu	32,7 %	33,0 %	32,0 %
	trifft teilweise zu	26,2 %	20,4 %	16,5 %
	trifft ziemlich zu	13,1 %	16,5 %	15,5 %
	trifft voll und ganz zu	6,5 %	3,9 %	8,2 %
	kann ich nicht sagen	-	1,9 %	-
	keine Markierung	0,9 %	-	2,1 %
N		107	103	97
K (SE)		-0,53 (0,47)	-0,74 (0,48)	-0,80 (0,49)
S (SE)		0,46 (0,24)	0,45 (0,24)	0,54 (0,25)
Mean (SD)		2,52 (1,16)	2,42 (1,15)	2,47 (1,27)
Shapiro-Wilk		0,897***	0,886***	0,875***
N		106	101	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 14 Verteilung des Item 5 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	3,7 %	1,9 %	2,1 %
	trifft wenig zu	17,8 %	16,5 %	21,6 %
	trifft teilweise zu	44,9 %	38,8 %	35,1 %
	trifft ziemlich zu	29,9 %	22,3 %	21,6 %
	trifft voll und ganz zu	3,7 %	18,4 %	17,5 %
	kann ich nicht sagen	-	1,9 %	1,0 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-0,04 (0,46)	-0,73 (0,48)	-0,87 (0,49)
S (SE)		-0,24 (0,23)	0,07 (0,24)	0,12 (0,25)
Mean (SD)		3,12 (0,88)	3,40 (1,04)	3,32 (1,07)
Shapiro-Wilk		0,882***	0,892***	0,895***
N		107	101	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 15 Verteilung des Item 6 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	11,2 %	13,6 %	8,2 %
	trifft wenig zu	45,8 %	34,0 %	44,3 %
	trifft teilweise zu	35,5 %	29,1 %	26,8 %
	trifft ziemlich zu	5,6 %	17,5 %	15,5 %
	trifft voll und ganz zu	1,9 %	4,9 %	3,1 %
	kann ich nicht sagen	-	1,0 %	1,0 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		0,64 (0,46)	-0,57 (0,47)	-0,29 (0,49)
S (SE)		0,49 (0,23)	0,29 (0,24)	0,52 (0,25)
Mean (SD)		2,41 (0,84)	2,66 (1,08)	2,60 (0,96)
Shapiro-Wilk		0,861***	0,904***	0,878***
N		107	102	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 16 Verteilung des Item 7 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	13,1 %	8,7 %	10,3 %
	trifft wenig zu	29,0 %	26,2 %	33,0 %
	trifft teilweise zu	39,3 %	34,0 %	22,7 %
	trifft ziemlich zu	15,0 %	16,5 %	25,8 %
	trifft voll und ganz zu	3,7 %	14,6 %	6,2 %
	kann ich nicht sagen	-	-	1,0 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		-0,34 (0,46)	-0,74 (0,47)	-0,92 (0,49)
S (SE)		0,13 (0,23)	0,19 (0,24)	0,14 (0,25)
Mean (SD)		2,67 (1,01)	3,02 (1,17)	2,84 (1,12)
Shapiro-Wilk		0,905***	0,907***	0,905***
N		107	103	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 17 Verteilung des Item 8 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	15,0 %	17,5 %	12,4 %
	trifft wenig zu	50,5 %	41,7 %	52,6 %
	trifft teilweise zu	30,8 %	29,1 %	21,6 %
	trifft ziemlich zu	1,9 %	9,7 %	7,2 %
	trifft voll und ganz zu	0,9 %	1,9 %	3,1 %
	kann ich nicht sagen	-	-	2,1 %
	keine Markierung	-	-	1,0 %
	doppelte Nennung	0,9 %	-	-
N		107	103	97
K (SE)		0,77 (0,47)	-0,08 (0,47)	1,02(0,49)
S (SE)		0,39 (0,24)	0,46 (0,24)	0,93 (0,25)
Mean (SD)		2,23 (0,76)	2,37 (0,95)	2,34 (0,91)
Shapiro-Wilk		0,836***	0,886***	0,838***
N		106	103	94

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 18 Verteilung des Item 9 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	15,9 %	13,6 %	5,2 %
	trifft wenig zu	36,4 %	26,2 %	37,1 %
	trifft teilweise zu	28,0 %	35,0 %	33,0 %
	trifft ziemlich zu	15,0 %	20,4 %	17,5 %
	trifft voll und ganz zu	2,8 %	4,9 %	5,2 %
	kann ich nicht sagen	-	-	1,0 %
	keine Markierung	1,9 %	-	1,0 %
N		107	103	97
K (SE)		-0,49 (0,47)	-0,63 (0,47)	-0,33 (0,49)
S (SE)		0,34 (0,24)	0,05 (0,24)	0,42 (0,25)
Mean (SD)		2,51 (1,03)	2,77 (1,08)	2,80 (0,97)
Shapiro-Wilk		0,899***	0,913***	0,889***
N		105	103	95

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 19 Verteilung des Item 10 des LIST-Inventars

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	trifft überhaupt nicht zu	17,8 %	14,6 %	12,4 %
	trifft wenig zu	46,7 %	38,8 %	46,4 %
	trifft teilweise zu	24,3 %	21,4 %	22,7 %
	trifft ziemlich zu	7,5 %	19,4 %	9,3 %
	trifft voll und ganz zu	3,7 %	4,9 %	6,2 %
	kann ich nicht sagen	-	1,0 %	2,1 %
	keine Markierung	-	-	1,0 %
N		107	103	97
K (SE)		0,52 (0,46)	-0,70 (0,47)	0,26 (0,49)
S (SE)		0,78 (0,23)	0,39 (0,24)	0,81 (0,25)
Mean (SD)		2,33 (0,98)	2,61 (1,11)	2,49 (1,05)
Shapiro-Wilk		0,864***	0,888***	0,861***
N		107	102	94

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 20 Deskriptive Statistik des summierten Index der LIST-Items

	NO-DK	Layout SEP-DK	ADD-DK
K (SE)	0,03 (0,47)	0,34 (0,49)	-0,60 (0,50)
S (SE)	0,03 (0,24)	0,09 (0,25)	-0,09 (0,25)
Mean (SD)	2,58 (,54)	2,74 (0,62)	2,68 (0,57)
Shapiro-Wilk	0,986	0,984	0,987
N	105	97	92

Tabelle A 21 Ergebnisse der Varianzanalyse der LIST-Items

	NO-DK	Layout SEP-DK	ADD-DK	
	Mean (SD)	Mean (SD)	Mean (SD)	F-Test
Item 1	2,71 (1,12)	2,82 (1,06)	2,63 (1,16)	$F_{(2, 300)} = 0,72$ $p = 0,486$
Item 2	3,10 (1,16)	2,96 (1,23)	3,04 (1,30)	$F_{(2, 302)} = 0,35$ $p = 0,703$
Item 3	2,19 (1,06)	2,26 (1,03)	2,16 (1,14)	$F_{(2, 301)} = 0,26$ $p = 0,769$
Item 4	2,52 (1,16)	2,42 (1,15)	2,47 (1,27)	$F_{(2, 299)} = 0,19$ $p = ,824$
Item 5	3,12 (0,88)	3,40 (1,04)	3,32 (1,07)	$F_{(2, 300)} = 2,09$ $p = 0,125$
Item 6	2,41 (0,84)	2,66 (1,08)	2,60 (0,96)	$F_{(2, 301)} = 1,88$ $p = 0,155$
Item 7	2,67 (1,01)	3,02 (1,17)	2,84 (1,12)	$F_{(2, 302)} = 2,60$ $p = 0,076$
Item 8	2,23 (0,76)	2,37 (0,95)	2,34 (0,91)	$F_{(2, 300)} = 0,77$ $p = 0,462$
Item 9	2,51 (1,03)	2,77 (1,08)	2,80 (0,97)	$F_{(2, 300)} = 2,36$ $p = 0,096$
Item 10	2,33 (0,98)	2,61 (1,11)	2,49 (1,05)	$F_{(2, 300)} = 1,91$ $p = 0,151$
Summierter LIST-Index	2,58 (0,54)	2,74 (0,62)	2,68 (0,57)	$F_{(2, 291)} = 2,01$ $p = 0,136$

Tabelle A 22 Verteilung des Items Sicherheit zu den PES-Items nach Skalenlayout

	NO-DK	Layout SEP-DK	ADD-DK
überhaupt nicht sicher	1,9 %	-	1,1 %
weniger sicher	4,7 %	5,9 %	3,2 %
mittelmäßig sicher	21,7 %	29,7 %	31,9 %
ziemlich sicher	50,0 %	56,4 %	50,0 %
sehr sicher	21,7 %	7,9 %	13,8 %
N	106	101	94

 $\chi^2 = 12,09; p > 0,05$
Tabelle A 23 Verteilung des Items Wichtigkeit des Themas der PES-Items nach Skalenlayout

	NO-DK	Layout SEP-DK	ADD-DK
überhaupt nicht wichtig	-	1,0 %	1,0 %
weniger wichtig	1,9 %	3,9 %	4,2 %
mittelmäßig wichtig	27,1 %	17,5 %	22,9 %
ziemlich wichtig	42,1 %	49,5 %	49,0 %
sehr wichtig	29,0 %	28,2 %	22,9 %
N	107	103	96

 $\chi^2 = 5,81; p > 0,05$

Tabelle A 24 Verteilung des Items Sicherheit zu den LIST-Items nach Skalenlayout

	NO-DK	Layout SEP-DK	ADD-DK
überhaupt nicht sicher	0,9	-	1,0 %
weniger sicher	1,9 %	1,0 %	2,1 %
mittelmäßig sicher	11,2 %	14,6 %	7,3 %
ziemlich sicher	57,9 %	57,3 %	69,8 %
sehr sicher	28,0 %	27,2 %	19,8 %
N	107	103	96

$\chi^2 = 7,03; p > 0,05$

Tabelle A 25 Verteilung des Items Wichtigkeit des Themas der LIST-Items nach Skalenlayout

	NO-DK	Layout SEP-DK	ADD-DK
überhaupt nicht wichtig	0,0 %	0,0 %	5,3 %
weniger wichtig	7,5 %	7,8 %	6,3 %
mittelmäßig wichtig	22,4 %	23,5 %	25,3 %
ziemlich wichtig	48,6 %	49,0 %	48,4 %
sehr wichtig	21,5 %	19,6 %	14,7 %
N	107	102	95

$\chi^2 = 12,63; p > 0,05$

Tabelle A 26 Ergebnisse der Varianzanalyse der Items Sicherheit und Wichtigkeit zu den Themen der PES und des LIST-Inventars nach Skalenlayout

	NO-DK	Layout SEP-DK	ADD-DK	
	Mean (SD)	Mean (SD)	Mean (SD)	F-Test
PES-Items				
Item Sicherheit	3,85 (0,88)	3,66 (0,71)	3,72 (0,78)	$F_{(2, 298)} = 1,47$ $p = 0,232$
Item Wichtigkeit	3,98 (0,80)	4,00 (0,84)	3,89 (0,84)	$F_{(2,303)} = 0,54$ $p = 0,581$
Multiplikativer Index Einstellungsstärke	15,51 (5,30)	14,74 (4,45)	14,68 (5,02)	$F_{(2, 298)} = 0,90$ $p = 0,408$
LIST-Items				
Item Sicherheit	4,10 (0,74)	4,11 (0,67)	4,05 (0,67)	$F_{(2, 303)} = 0,19$ $p = 0,827$
Item Wichtigkeit	3,84 (0,85)	3,80 (0,85)	3,61 (1,00)	$F_{(2,301)} = 1,89$ $p = 0,153$
Multiplikativer Index Einstellungsstärke	15,93 (5,26)	15,75 (4,85)	14,68 (5,06)	$F_{(2, 301)} = 1,75$ $p = 0,176$

Tabelle A 27 Verteilung der dichotomen Variable der Einstellungsstärke zu den Inhalten der PES und des LIST-Inventars nach Skalenlayout

Einstellungsstärke gegenüber	NO-DK	Layout SEP-DK	ADD-DK
Pol. Selbstwirksamkeit (PES)			
- niedrig-mittel	44,3 %	48,5 %	48,9 %
- hoch	55,7 %	51,5 %	51,1 %
N	106	101	94
Lernstrategien (LIST)			
- niedrig-mittel	40,2 %	41,2 %	38,9 %
- hoch	59,8 %	58,8 %	61, %
N	107	102	95

Politische Selbstwirksamkeit: $\chi^2 = 0,532; p > 0,05$

Lernstrategien: $\chi^2 = 0,102; p > 0,0$

Tabelle A 28 Ergebnisse der Varianzanalyse der EU-Items

	NO-DK	Layout SEP-DK	ADD-DK	
	Mean (SD)	Mean (SD)	Mean (SD)	F-Test
Item 1	3,17 (0,98)	3,08 (1,06)	3,35 (1,12)	$F_{(2, 435)} = 2,365$ $p = 0,095$
Item 2	3,08 (0,93)	3,82 (1,00)	4,04 (1,00)	$F_{(2, 474)} = 2,615$ $p = 0,074$
Item 3	3,61 (1,13)	3,81 (1,14)	3,97 (1,13)	$F_{(2, 472)} = 3,788$ $p = 0,023$
Item 4	2,93 (0,99)	3,22 (1,03)	3,38 (1,10)	$F_{(2, 443)} = 7,257$ $p = 0,001$
Item 5	3,22 (1,11)	3,00 (1,09)	3,26 (1,10)	$F_{(2, 456)} = 2,304$ $p = 0,101$
Item 6	3,46 (1,29)	3,59 (1,26)	3,73 (1,31)	$F_{(2, 473)} = 1,747$ $p = 0,175$
Item 7	3,38 (1,14)	3,25 (1,13)	3,53 (1,10)	$F_{(2, 450)} = 2,295$ $p = 0,102$
Item 8	2,95 (1,22)	3,17 (1,22)	3,01 (1,34)	$F_{(2, 471)} = 1,229$ $p = 0,293$
Item 9	2,91 (1,25)	3,11 (1,34)	3,25 (1,32)	$F_{(2, 466)} = 2,473$ $p = 0,085$
Item 10	3,88 (0,89)	4,04 (0,95)	4,24 (0,88)	$F_{(2, 467)} = 6,104$ $p = 0,002$
Summierter Index EU-Items	3,32 (0,37)	3,43 (0,38)	3,57 (0,38)	$F_{(2, 351)} = 13,450$ $p = 0,000$

Tabelle A 29 Verteilung des EU-Item 1

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	2,5 %	5,2 %	5,0 %
	lehne eher ab	25,3 %	21,8 %	14,9 %
	weder noch	31,0 %	21,8 %	25,5 %
	stimme eher zu	32,3 %	27,6 %	26,7 %
	stimme stark zu	7,6 %	5,7 %	14,3 %
	weiß nicht	-	16,7 %	12,4 %
	keine Markierung	1,3 %	1,1 %	1,2 %
N		158	174	161
K (SE)		-0,78 (0,40)	-0,96 (0,45)	-0,72 (0,49)
S (SE)		-0,02 (0,20)	-0,03 (0,23)	-0,19 (0,25)
Shapiro-Wilk		0,896***	0,900***	0,906***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 30 Verteilung des EU-Item 2

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	1,3 %	1,7 %	2,5 %
	lehne eher ab	11,4 %	12,6 %	9,3 %
	weder noch	10,8 %	9,2 %	3,1 %
	stimme eher zu	55,7 %	50,0 %	47,8 %
	stimme stark zu	19,6 %	22,4 %	32,9 %
	weiß nicht	-	3,4 %	3,1 %
	keine Markierung	1,3 %	0,6 %	1,2 %
N		158	174	161
K (SE)		0,24 (0,40)	-0,63 (0,45)	0,58 (0,49)
S (SE)		-0,84 (0,20)	-0,64 (0,23)	-1,08 (0,25)
Shapiro-Wilk		0,824***	0,832***	0,791***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 31 Verteilung des EU-Item 3

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	5,1 %	4,6 %	1,9 %
	lehne eher ab	17,1 %	14,4 %	14,3 %
	weder noch	8,2 %	4,0 %	6,2 %
	stimme eher zu	49,4 %	47,1 %	32,3 %
	stimme stark zu	19,0 %	28,2 %	37,3 %
	weiß nicht	-	1,1 %	2,5
	keine Markierung	1,3 %	0,6 %	5,6
N		158	174	161
K (SE)		-0,39 (0,40)	0,02 (0,45)	-0,29 (0,49)
S (SE)		-0,76 (0,20)	-1,00 (0,23)	-0,94 (0,25)
Shapiro-Wilk		0,832***	0,796***	0,790***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 32 Verteilung des EU-Item 4

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	7,0 %	3,4 %	4,3 %
	lehne eher ab	27,8 %	23,0 %	16,8 %
	weder noch	32,3 %	17,2 %	18,0 %
	stimme eher zu	28,5 %	38,5 %	34,8 %
	stimme stark zu	3,2 %	5,2 %	11,8 %
	weiß nicht	-	12,1 %	12,4 %
	keine Markierung	1,3 %	0,6 %	1,9 %
N		158	174	161
K (SE)		-0,69 (0,40)	-0,48 (0,45)	-0,51 (0,49)
S (SE)		-0,09 (0,20)	-0,58 (0,23)	-0,55 (0,25)
Shapiro-Wilk		0,899***	0,853***	0,876***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 33 Verteilung des EU-Item 5

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	5,1 %	4,0 %	3,7 %
	lehne eher ab	26,6 %	32,8 %	24,2 %
	weder noch	17,1 %	23,6 %	18,0 %
	stimme eher zu	38,6 %	23,0 %	33,5 %
	stimme stark zu	9,5 %	9,2 %	10,6 %
	weiß nicht	-	5,2 %	5,6 %
	keine Markierung	3,2 %	2,3 %	4,3 %
N		158	174	161
K (SE)		-1,05 (0,40)	-0,94 (0,45)	-1,05 (0,49)
S (SE)		-0,18 (0,20)	0,26 (0,23)	0,06 (0,25)
Shapiro-Wilk		0,879***	0,887***	0,895***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 34 Verteilung des EU-Item 6

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	7,0 %	4,6 %	5,0 %
	lehne eher ab	24,1 %	24,7 %	20,5 %
	weder noch	6,3 %	3,4 %	6,2 %
	stimme eher zu	36,7 %	38,5 %	26,7 %
	stimme stark zu	22,8 %	26,4 %	36,6 %
	weiß nicht	-	1,1 %	1,2 %
	keine Markierung	3,2 %	1,1 %	3,7 %
N		158	174	161
K (SE)		-1,08 (0,40)	-1,18 (0,45)	-0,64 (0,49)
S (SE)		-0,46 (0,20)	-0,51 (0,23)	-0,83 (0,25)
Shapiro-Wilk		0,853***	0,821***	0,805***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 35 Verteilung des EU-Item 7

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	5,1 %	6,3 %	2,5 %
	lehne eher ab	19,6 %	19,0 %	18,0 %
	weder noch	22,2 %	22,4 %	14,9 %
	stimme eher zu	32,9 %	31,6 %	36,6 %
	stimme stark zu	16,5 %	11,5 %	16,8 %
	weiß nicht	-	8,0 %	8,1 %
	keine Markierung	3,8 %	1,1 %	3,1 %
N		158	174	161
K (SE)		-0,89 (0,40)	-0,54 (0,45)	-1,14 (0,49)
S (SE)		-0,23 (0,20)	-0,40 (0,23)	-0,21 (0,25)
Shapiro-Wilk		0,904***	0,901***	0,879***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 36 Verteilung des EU-Item 8

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	13,3 %	9,2 %	16,8 %
	lehne eher ab	25,3 %	24,1 %	22,4 %
	weder noch	19,0 %	16,1 %	13,7 %
	stimme eher zu	31,0 %	34,5 %	28,6 %
	stimme stark zu	8,2 %	12,1 %	14,3 %
	weiß nicht	-	3,4 %	1,9 %
	keine Markierung	3,2 %	0,6 %	2,5 %
N		158	174	161
K (SE)		-1,07 (0,40)	-0,93 (0,45)	-0,41 (0,49)
S (SE)		-0,07 (0,20)	-0,35 (0,23)	-0,65 (0,25)
Shapiro-Wilk		0,901***	0,888***	0,875***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 37 Verteilung des EU-Item 9

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	14,6 %	14,9 %	9,9 %
	lehne eher ab	27,8 %	21,3 %	22,4 %
	weder noch	14,6 %	11,5 %	15,5 %
	stimme eher zu	32,3 %	32,8 %	25,5 %
	stimme stark zu	8,2 %	14,4 %	19,9 %
	weiß nicht	-	4,6 %	4,3 %
	keine Markierung	2,5 %	0,6 %	2,5 %
N		158	174	161
K (SE)		-1,22 (0,40)	-1,29 (0,45)	-1,27 (0,49)
S (SE)		-0,03 (0,20)	-0,21 (0,23)	-0,17 (0,25)
Shapiro-Wilk		0,886***	0,871***	0,884***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 38 Verteilung des EU-Item 10

		NO-DK	Layout SEP-DK	ADD-DK
Antwortkategorie	lehne stark ab	0,,6 %	2,9 %	1,2 %
	lehne eher ab	10,1 %	6,3 %	5,0 %
	weder noch	10,8 %	4,6 %	5,0 %
	stimme eher zu	54,4 %	52,3 %	41,6 %
	stimme stark zu	20,9 %	29,3 %	41,0 %
	weiß nicht	-	4,0 %	3,1 %
	keine Markierung	3,2 %	0,6 %	3,1 %
N		158	174	161
K (SE)		0,44 (0,40)	1,37 (0,45)	2,42 (0,49)
S (SE)		-0,88 (0,20)	-1,23 (0,23)	-1,47 (0,25)
Shapiro-Wilk		0,808***	0,777***	0,742***
N		144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 39 Deskriptive Statistik des summierten Index der EU-Items

	NO-DK	Layout SEP-DK	ADD-DK
K (SE)	1,58 (0,40)	2,23 (0,45)	-0,19 (0,49)
S (SE)	-0,50 (0,20)	-0,39 (0,23)	-0,18 (0,25)
Mean (SD)	3,32 (0,37)	3,43 (0,38)	3,57 (0,38)
Shapiro-Wilk	0,975**	0,965**	0,987
N	144	113	97

***p < 0,001; **p<0,01, * p<0,05

Tabelle A 40 Ergebnisse der logistischen Multilevel-Regression zur Wahl (a) der rechten Extremkategorie, (b) der Mittelkategorie und (c) der DK-Kategorie mit Interaktion „Bildung×Layout“

	(a) Wahl der rechten Extremkategorie β (se, robust)	(b) Wahl der Mittelkategorie β (se, robust)	(c) DK-Kategorie ⁺ β (se, robust)
Layout			
NO-DK	-0,502** (0,17)	0,228 (0,19)	-
SEP -DK	-0,152 (0,15)	-0,213 (0,19)	0,092 (0,26)
Geschlecht			
Mann	0,324*** (0,09)	-0,104 (0,09)	-0,699*** (0,16)
Bildung			
hoch	-0,0159 (0,13)	0,008 (0,16)	-0,174 (0,25)
Bildung×Layout			
hoch×NO-DK	-0,301 (0,21)	0,184 (0,23)	-
hoch×SEP-DK	-0,545 (0,19)	0,353 (0,23)	0,013 (0,33)
Altersgruppe			
Ältere (55 Jahre +)	0,377*** (0,09)	-0,261* (0,11)	-0,985*** (0,25)
Konstante	-2,237*** (0,20)	-0,997*** (0,15)	-1,186*** (0,25)
McFadden's-R ²	0,085	0,062	0,104
AIC	3950,93	3595,14	1271,69
BIC	4059,98	3704,19	1362,32
N	4512	4512	3109

Referenzkategorie: ADD-DK-Layout; Frau, niedrige/mittlere Bildung, Jüngere bis einschließlich 54 Jahre

* p<0,05, ** p<0,01, *** p<0,001

⁺ Von der logistischen Multilevel-Regression der Wahl der DK-Kategorie wurden alle Befragte der Layoutgruppe „NO-DK“ ausgeschlossen.

Tabelle A 41 Ergebnisse der logistischen Multilevel-Regression zur Wahl (a) der rechten Extremkategorie, (b) der Mittelkategorie und (c) der DK-Kategorie mit Interaktion „Altersgruppe×Layout“

	(a) Wahl der rechten Extremkategorie β (se, robust)	(b) Wahl der Mittelkategorie β (se, robust)	(c) DK-Kategorie ⁺ β (se, robust)
Layout			
NO-DK	-0,571*** (0,12)	0,308** (0,12)	-
SEP -DK	-0,452*** (0,11)	0,047 (0,12)	0,190 (0,17)
Geschlecht			
Mann	0,323*** (0,08)	-0,107 (0,09)	-0,709*** (0,16)
Bildung			
hoch	-0,243** (0,08)	0,187* (0,09)	-0,163 (0,17)
Altersgruppe			
Ältere (55 Jahre +)	0,513*** (0,14)	-0,295 (0,11)	-0,607 (0,33)
Altersgruppe×Layout			
Ältere×NO-DK	-0,403 (0,23)	0,188 (0,25)	-
Ältere×SEP-DK	-0,11 (0,21)	-0,101 (0,27)	-0,807 (0,52)
Konstante	-2,119*** (0,19)	-1,105*** (0,15)	-1,124*** (0,18)
McFadden's-R ²	0,084	0,062	0,106
AIC	3955,72	3596,13	1269,14
BIC	4064,7723	3705,18	1359,77
N	4512	4512	3109

Referenzkategorie: ADD-DK-Layout; Frau, niedrige/mittlere Bildung, Jüngere bis einschließlich 54 Jahre

* p<0,05, ** p<0,01, *** p<0,001

⁺ Von der logistischen Multilevel-Regression der Wahl der DK-Kategorie wurden alle Befragte der Layoutgruppe „NO-DK“ ausgeschlossen.